

# Применение методов машинного обучения для прогнозирования опасных отказов объектов железнодорожного пути

Игорь Б. Шубинский<sup>1</sup>, Алексей М. Замышляев<sup>1</sup>, Ольга Б. Проневич<sup>1</sup>, Алексей Н. Игнатов<sup>2</sup>, Евгений Н. Платонов<sup>2\*</sup>

<sup>1</sup>АО «НИИАС», Российская Федерация, Москва, <sup>2</sup>Московский авиационный институт, Российская Федерация, Москва

\*en.platonov@gmail.com



Игорь Б.  
Шубинский



Алексей М.  
Замышляев



Ольга Б. Проневич



Алексей Н.  
Игнатов



Евгений Н.  
Платонов

**Резюме.** Цель работы – уменьшение количества опасных событий на железнодорожном пути за счет разработки методологии прогнозирования редких опасных отказов на основе обработки больших массивов данных, поступающих в реальном масштабе времени от диагностических систем о каждом километре пути. Опасные отказы являются редкими событиями. Однако для железной дороги в целом ежегодно количество таких событий измеряется десятками и для эффективного управления безопасностью движения важно перейти от оценки вероятности появления опасного отказа как такового к определению наиболее вероятного места отказа. **Методы.** Задача выявления редких, но опасных возможных событий из сотен тысяч данных о не критичных отклонениях параметров железнодорожного пути не поддается решению традиционными методами статистической обработки. Прогнозирование опасных событий на основе указанной статистики осуществляется с помощью методов искусственного интеллекта. Для этого используются технологии анализа больших данных и Data Science. К таким технологиям относятся методы машинного обучения, решающие задачи классификации объектов на основе его характеристик (признаков, предикатов) и известных случаях реализации нежелательных событий. Применение различных алгоритмов машинного обучения демонстрируется на примере прогнозирования отказов верхнего строения пути по наблюдениям, собранным в период с 2014 г. по 2019 г. на Куйбышевской железной дороге. **Результаты/Выводы.** Результатом категорирования объектов является вывод о наиболее вероятном месте появления опасного отказа на железнодорожном пути. Такой вывод основан на анализе связей между случаями появления нежелательных событий и фактическими характеристиками объекта и условиями его эксплуатации. Практическое значение этой работы состоит в том, что предлагаемый набор методов и средств может рассматриваться как неотъемлемая часть системы принятия решений по техническому обслуживанию пути. Его можно легко адаптировать для функционирования в режиме онлайн и интегрировать с автоматизированной измерительной системой на базе транспортного средства.

**Ключевые слова:** машинное обучение, отказы объектов железнодорожного пути, решающие деревья.

**Для цитирования:** Шубинский И.Б., Замышляев А.М., Проневич О.Б., Платонов Е.Н., Игнатов А.Н. Применение методов машинного обучения для прогнозирования опасных отказов объектов железнодорожного пути // Надежность. 2020. № 2. С 43-53. <https://doi.org/10.21683/1729-2646-2020-20-2-43-53>

Поступила 29.02.2020 г. / После доработки 18.04.2020 г. / К печати 17.06.2020 г.

## 1. Введение

С каждым годом увеличивается роль цифровых технологий в управлении технологическими предприятиями. Автоматизированные системы управления (АСУ) позволяют существенно увеличить скорость выполнения операций на производстве, системы автопилотирования используются в поездах и самолетах, обеспечивая безопасность движения на скоростях, выходящих за грани человеческой реакции. Современные инструменты диагностики замечают то, что не способен увидеть человеческий глаз, и применяются в медицине, машиностроении, исследовании космоса и других областях науки и промышленности. Но цифровой мир не ограничивается автоматизацией процессов, которые не может выполнить человек, особенно, если дело касается крупных предприятий. Компания ОАО «РЖД» с 2016 года выстроила систему электронного документооборота более чем с тысячей предприятий, вовлеченных в процесс грузовых перевозок [1]. Только об электропоездах «Ласточка» диагностическая информация собирается с помощью 342 датчиков и измерительных устройств. В ОАО «РЖД» одновременно с системой диагностики состояния локомотивов функционируют десятки АСУ, обеспечивающих холдинг информации о состоянии железнодорожного пути [2, 3], оборудования железнодорожного электроснабжения [4], безопасности движения [5], о графиках исполнения движения [6] и множестве других объектов и процессов. Каждая из АСУ ОАО «РЖД» призвана решать отдельные задачи, но для комплексного управления железнодорожным транспортом были разработаны системы корпоративного уровня ЕКП УРРАН (Единая корпоративная платформа Управления Ресурсами, Рисками и Надежностью на стадиях жизненного цикла),

ЕК АСУИ (Единая корпоративная автоматизированная система управления инфраструктурой), ЕК АСУТР (Единая Корпоративная Автоматизированная Система Управления Трудовыми Ресурсами), ЕК АСУФР (Единая корпоративная автоматизированная система управления финансами и ресурсами). Существующие системы сбора и хранения данных, а также корпоративные системы, агрегирующие информацию из различных источников, обеспечивают ОАО «РЖД» возможность успешного применения технологий обработки и управления большими данными (Big Data) Data Science (рис. 1).

## 2. Актуальность прогнозирования появления опасных отказов верхнего строения пути

Высокая интенсивность движения поездов, высокая скорость движения на железнодорожной сети, условия окружающей среды, старение – все это вызывает износ железных дорог, в первую очередь – железнодорожного пути. Дефекты рельсов могут привести к сходу, аварии или крушению поезда. Такие опасные события сопровождаются повреждением железнодорожных путей, электросетей, а также вагонов и секций локомотив вплоть до их исключения из инвентарного парка [7]. Также может произойти выход сошедших подвижных единиц в габарит соседнего пути, что может привести к столкновению со встречным поездом и, как следствие, привести к катастрофическому ущербу от схода поезда [8, 9]. Большая доля нежелательных событий из-за состояния пути характерна не только для железных дорог России. За последнее десятилетие около трети всех железнодорожных происшествий в США были вызваны дефектами, связанными с железнодорожными путями [10].



Рис. 1. АСУ ОАО «РЖД» – основа для применения технологии анализа Big Data

При анализе сходов, аварий и крушений единиц подвижного состава грузового поезда было выявлено, что эти события, вызванные неисправностью железнодорожного пути, могли произойти на километре пути, оцененном, в частности, на «хорошо». В этой связи для прогноза состояния пути недостаточно агрегированной оценки состояния данного километра, здесь необходимо принимать в расчет другие параметры: количество уширений, рихтовок и др. Однако одного сбора дополнительных параметров недостаточно. Согласно исследованию [11], лишь часть данных о контролируемом объекте является полезной с точки зрения принятия решения при управлении конкретными событиями (рис. 2).

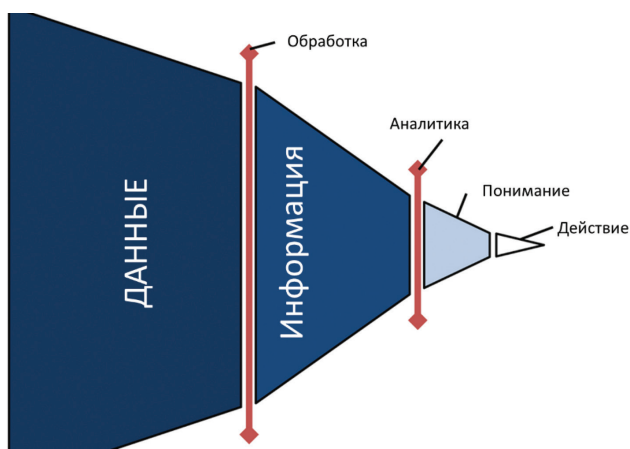


Рис. 2. Преобразование больших объемов необработанных данных в практическую информацию

Современные методы многофакторного анализа данных и технологии машинного обучения обеспечивают возможность включать в модели более 50 факторов. Они позволяют на основе текущих значений измеренных признаков, характеризующих состояние железнодорожного пути, делать вывод о необходимости и месте проведения срочных ремонтных работ, чтобы предотвратить отказы железнодорожного пути и сходы, аварии и крушения поездов. Об эффективности применения технологии анализа Big Data и Data Science можно говорить на основе имеющегося практического международного опыта, анализ которого приведен ниже.

### 3. Обзор методов машинного обучения и их применение в анализе дефектов железнодорожного пути

В настоящее время из-за быстрого технического прогресса и повсеместного развертывания недорогих датчиков, широкое использование беспроводных средств связи привело к повышению роли интернет-технологий в эффективной реализации стратегий технического обслуживания во многих отраслях. В железнодорожной отрасли также активно используются технологии Data Science [12].

За последние два десятилетия машинное обучение произвело революцию в широком спектре областей, таких, как компьютерное зрение, обработка естественного языка и распознавание речи. С ростом количества данных, собираемых устройствами мониторинга, такими, как беспроводные сенсорные сети или видеокамеры высокого разрешения, которые широко используются для проверки критически важной железнодорожной инфраструктуры, машинное обучение также приобретает все большую популярность для улучшения работы и повышения надежности железнодорожных систем, а также позволяет минимизировать ежедневные затраты на обслуживание [13].

Методы машинного обучения можно разделить на классические методы обучения [14] и методы глубокого обучения [15]. Основное различие между этими методами заключается в уровне их представления. К классическим методам обучения относятся метод главных компонент, метод опорных векторов [16], деревья решений [17], случайный лес [18–20], логистическая регрессия [21] и метод ближайших соседей [22].

В работе [23] представлена методология систематизации данных для мониторинга состояния рельса. Основным внимание авторы направили на выявление закономерностей появления отказов на кривых малого радиуса, используя метод главных компонент и данные, полученные в результате ручного осмотра от шведской железнодорожной сети.

В работе [24] использован метод опорных векторов для прогнозирования ситуации, когда не очень серьезные дефекты пути перерастают в более серьезные типы дефектов и отказы.

В работе [25] на основе решающих деревьев разработана система предварительного автоматического ранжирования инцидентов, которая по имеющимся признакам оценивает вероятность предотказного состояния объекта.

Цзян и соавторы [26] предложили гибридный подход к распознаванию усталости при качении по данным, полученным с помощью ультразвукового лазера.

В работе [10] метод главных компонент применен совместно с методом опорных векторов к набору данных из 31 объектов, собранных для железнодорожной сети I класса США для обнаружения четырех видов поверхностных дефектов.

В последнее время научное сообщество также использовало преимущества методов глубокого обучения для изучения дефектов рельсов. Исследователи полагают, что глубокое обучение может стать элементом полностью автоматизированных систем железнодорожного мониторинга [27].

Алгоритмы глубокого обучения, основанные на нейронных сетях, применяются в качестве основного инструмента для выявления структурных дефектов рельса. Наиболее широко используются сверточные нейронные сети (CNN). Это произошло благодаря широкому использованию видеокамер, которые обеспечивают



исследовательское сообщество огромным количеством данных и дают возможность использовать более сложные методы обучения. Однако CNN являются «черным ящиком» и практически не интерпретируются. Другими словами, исследователь машинного обучения не может объяснить, как модель CNN придумала свои прогнозы, или доказать их надежность для конечного пользователя.

В работе [28] на основе технологии CNN рассматриваются подходы к решению задачи автоматизированной обработки изображений верхнего строения железнодорожного пути (ВСП) с целью выявления зон, подозрительных на дефекты. За основу были взяты изображения, полученные одним из поездов Центра диагностики и мониторинга устройств инфраструктуры – структурного подразделения Западно-Сибирской железной дороги.

Ли и соавторы [29] использовали искусственные нейронные сети и метод опорных векторов для прогнозирования износа балластной призмы на основе таких факторов как кривизна, пропущенный тоннаж и др. При этом авторы заметили, что для получения стабильных прогнозов необходимо проводить измерения минимум два года.

Более подробный обзор применения различных методов машинного обучения к выявлению дефектов железнодорожного пути можно найти в обзоре [30].

Многообразие используемых моделей свидетельствует о том, что применение технологий машинного обучения в настоящее время является исследовательским процессом, включающим в себя этапы:

- анализ источников информации о состоянии железнодорожного пути;
- подготовка данных к машинному обучению;
- постановка задачи машинного обучения;
- обучение моделей;
- выбор лучшей модели;
- применение модели.

#### 4. Алгоритм подготовки данных о состоянии железнодорожного пути для применения методов машинного обучения в ОАО «РЖД»

Для подготовки данных, получаемых из АСУ ОАО «РЖД» разработан алгоритм, включающий в себя 5 этапов, приведенных в табл. 1.

Одним из основных понятий машинного обучения является выборка. Выборка – конечный набор прецедентов (случаев, событий, испытуемых образцов и др.) и соответствующие им данные (характеристики объектов), образующие описание прецедента. Выборка, включающая в себя полный набор доступных данных, должна содержать в себе целевую переменную – показатель, прогноз значения которого и является основной целью применения методов машинного обучения. Кроме этого, выборка делится на две части: обучающую выборку и тестовую выборку. Алгоритм подготовки данных, полученных из АСУ ОАО «РЖД», для формирования выборок, используемых в машинном обучении приведен на рис. 3.

Таблица 1 – Этапы подготовки данных

Наименование этапа	Цель	Условия выполнения этапа	Критерии необходимости этапа
Очистка данных	Повышение качества моделирования за счет улучшения качества данных	Выполняется всегда	Выполняется всегда
Преобразование данных	Повышение качества моделирования за счет сопоставления последовательностей с различными физическими единицами и/или диапазонами значений	Выполняется при необходимости для дискретных последовательностей	1. Диапазоны изменения значений различных признаков отличаются более чем в 5 раз. 2. Различные физические единицы измерения признаков
Дискретизация данных	Расширение диапазона доступных для применения моделей	Выполняется при необходимости для непрерывных последовательностей	1. Целевой признак является непрерывной величиной, но необходимо оценить вероятность попадания в интервал 2. Планируется применять метод, который не может использовать непрерывные данные.
Очистка текстов	Повышение качества моделирования за счет улучшения качества данных	Выполняется при необходимости для непрерывных последовательностей	Планируется использовать информацию из текстов при моделировании
Формирование выборок	Проверка качества разработанных моделей	Выполняется всегда	Выполняется всегда



## 5. Алгоритм применения методов машинного обучения для прогноза появления опасных отказов железнодорожного пути

Задачи машинного обучения обычно описываются в терминах того, как система машинного обучения должна обрабатывать обучающую выборку. В качестве прецедента обучающей выборки ВСП выбран километр ВСП, состояния которого характеризуются 77 параметрами, включая результаты диагностики, условия эксплуатации, качественные оценки. Значения этих параметров представляются в виде вектора  $x \in R^n$ , каждый элемент которого — значение признака.

Классификация — наиболее часто встречающаяся задача машинного обучения, она заключается в построении моделей, выполняющих отнесение интересующего нас объекта к одному из нескольких известных классов. В задачах этого типа алгоритм классификации должен ответить на вопрос, какой из категорий принадлежит объект. С точки зрения обеспечения безопасности движения каждый километр ВСП делится на два класса: 0 — километр с отсутствием опасного отказа ВСП, 1 — километр с опасным отказом ВСП.

На обучающей выборке подбирают наилучшие параметры для алгоритма классификации. На тестовой выборке вычисляют ошибку классификации для того, чтобы выбрать наилучший алгоритм.

Пусть  $X$  — пространство объектов, которые описываются набором признаков  $X = \{X^1, \dots, X^n\}^T$ ;  $Y = \{0, 1\}$  — множество допустимых откликов;  $y^*: X \rightarrow Y$  — целевая зависимость, известная только на объектах обучающей выборке  $Z^N = (x_i, y_i)_{i=1}^N$ , где  $x_i$  — вектор значений признаков, а  $y_i = y^*(x_i)$  — отклики целевой переменной,  $i = 1, \dots, N$ .

Задача обучения заключается в том, чтобы восстановить функциональную зависимость между объектами и откликами, то есть построить алгоритм  $a: X \rightarrow Y$ , аппроксимирующий целевую зависимость  $y^*(\cdot)$  на всем множестве  $X$ , а не только на объектах обучающей выборки  $Z^N$ .

На рис. 4 приведен алгоритм применения шести основных методов машинного обучения для классификации километра ВСП.

## 6. Критерии выбора лучшей модели

Для анализа точности алгоритма машинного обучения и сравнения точности алгоритмов существуют различные меры.

Для задачи бинарной классификации введем следующие обозначения:

TP — количество правильно спрогнозированных объектов с категорией «1»;

FN — количество объектов с истинной категорией «1», но прогнозом «0»;

FP — количество объектов с истинной категорией «0», но прогнозом «1»;

TN — количество правильно спрогнозированных объ-

ектов с категорией «0».

Основными мерами качества для моделей бинарной классификации являются.

1) Общая точность алгоритма  $AC = \frac{TP + TN}{TP + FP + FN + TN}$ , определяющая общую эффективность алгоритма давать правильные ответы.

2) Ложная тревога  $FPR = \frac{FP}{FP + TN}$ , показывающая возможность алгоритма прогнозировать ложные отклонения от нормального состояния.

3) Точность алгоритма  $PR = \frac{TP}{TP + FP}$ , показывающая, какая доля объектов, распознанных как объекты с категорией «1», предсказана верно.

4) Полнота алгоритма  $RE = \frac{TP}{TP + FN}$ , показывающая, какая доля объектов действительно имеющих категорию «1», предсказана верно.

5)  $F$ -мера алгоритма  $F = \frac{2 \cdot PR \cdot RE}{PR + RE}$  — среднее гармоническое точности и полноты.

6) Площадь под кривой ошибок AUC — глобальная характеристика качества, значения которой лежат в интервала от 0 до 1. Значение 0,5 соответствует случайному угадыванию, а значение 1 — безошибочному распознаванию. AUC является площадью под ROC-кривой. ROC-кривая показывает соотношение доли ложных положительных классификаций (FPR) и доли верных положительных классификаций (RE). ROC-кривая является достаточно сложной мерой точности алгоритма, более подробно с ней можно познакомиться в работе [31].

## 7. Численный эксперимент категорирования участков пути на основе прогноза появления отказа

Рассмотрим задачу классификации отказов для ВСП. Для предупреждения и предотвращения сходов, аварий и крушений состояние железнодорожного пути на сети железных дорог проверяется путеизмерителем на предмет отклонений от нормативных показателей. На основании полученных данных каждому километру железнодорожного пути присваивается определенная оценка: «неудовлетворительно», «удовлетворительно», «хорошо» и «отлично». Эти оценки должны сигнализировать об опасности возникновения транспортного происшествия, вызванного состоянием железнодорожного пути.

На Куйбышевской железной дороге в период с 2014 г. по 2019 г. собрана статистика по состоянию ВСП. Зарегистрированы отказы элементов железнодорожной инфраструктуры следующих видов: изолирующий стык, железобетонная шпала, рельсовая линия в целом, стыкование рельсов, геометрические параметры рельсовой



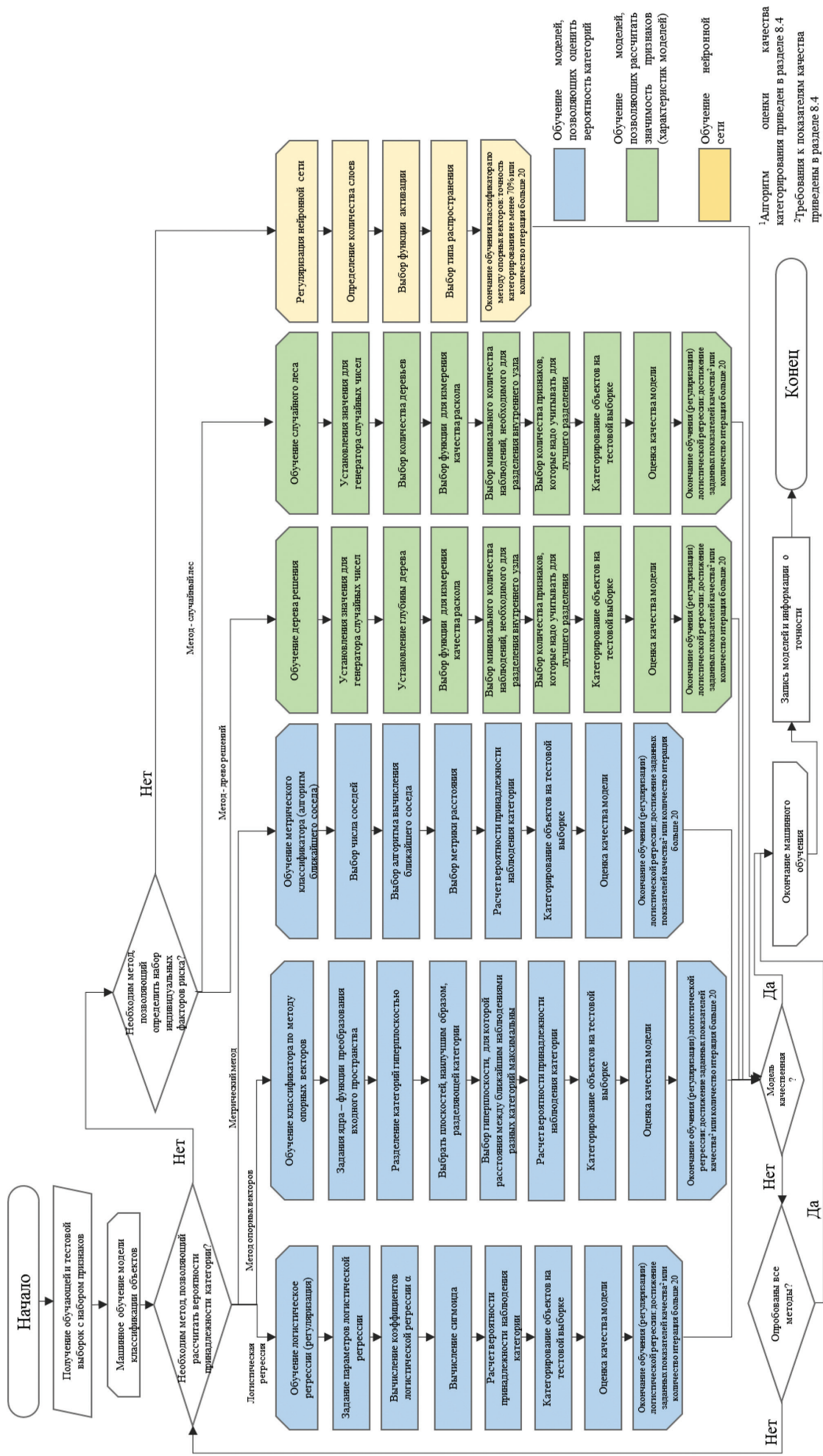


Рис. 4. Алгоритм применения методов машинного обучения для классификации состояния верхнего строения пути

Табл. 2. Показатели качества моделей

	Показатель качества	Логистическая регрессия (выборка № 2)	Решающее дерево (выборка № 2)	Случайный лес (выборка № 2)	Метод опорных векторов (выборка № 2)	Метод ближайших соседей (выборка № 1)
1	AC	0,74	0,76	0,75	0,73	0,72
2	FPR	0,41	<b>0,28</b>	<b>0,28</b>	0,41	0,46
3	PR	0,78	0,94	0,94	0,94	0,89
5	RE	0,78	0,94	0,94	0,94	0,89
6	F-мера	0,78	0,86	0,86	0,94	0,88
7	AUC	0,68	0,83	0,83	0,76	0,71

колеи и т.д. Для каждого километра пути за несколько лет были проведены ежемесячные измерения следующих показателей: количество уширений, количество отклонений, количество рихтовок, количество просадок, скорость движения на данном километре и т.д.

Если на километре железнодорожного пути зафиксирован отказ, то отклику присваивается значение «1», если нет, то значение «0», то есть множество категорий имеет вид  $Y=\{0,1\}$ . Требуется решить задачу бинарной классификации по наблюдениям за прошлые моменты времени и проверить работоспособность алгоритма на наблюдениях за 2019 г. На основании проведенной классификации прогнозируется появление опасного отказа.

Получено 194328 результатов наблюдений за различными объектами (километрами железнодорожного пути). Из них зафиксировано 267 объектов с опасными отказами. Данные разбиты на обучающую выборку (192375 объектов, из них 257 с опасными отказами, данные за 2014-2018 гг.) и тестовую выборку (1953 объекта, из них 10 с опасными отказами, данные за январь 2019 г.).

Для решения задачи классификации использовано несколько алгоритмов машинного обучения: логистическая

регрессия, алгоритм на основе решающих деревьев, метод случайного леса, метод опорных векторов (SVM) и метод ближайших соседей (NNK).

Сформированы обучающие выборки:

- обучающая выборка № 1: наблюдения за 2014-2018 годы на стандартизированных данных;
- обучающая выборка № 2: наблюдения за 2017-2018 годы на стандартизированных данных.

Дополнительно проведена процедура прореживания данных. Цель – повышение качества моделирования за счет формирования сбалансированных обучающих выборок, в которых количество наблюдений с категорией «1» должно составлять не менее 40% от общего числа наблюдений.

Для отбора признаков использован метод рекурсивного отбора признака для каждого метода машинного обучения.

На рис. 5 приведено сравнение качества моделей, а в табл. 2 приведены показатели качества моделей. В табл. 2 приведены модели, обученные на выборках, которые дали лучшие показатели качества для своего типа модели.

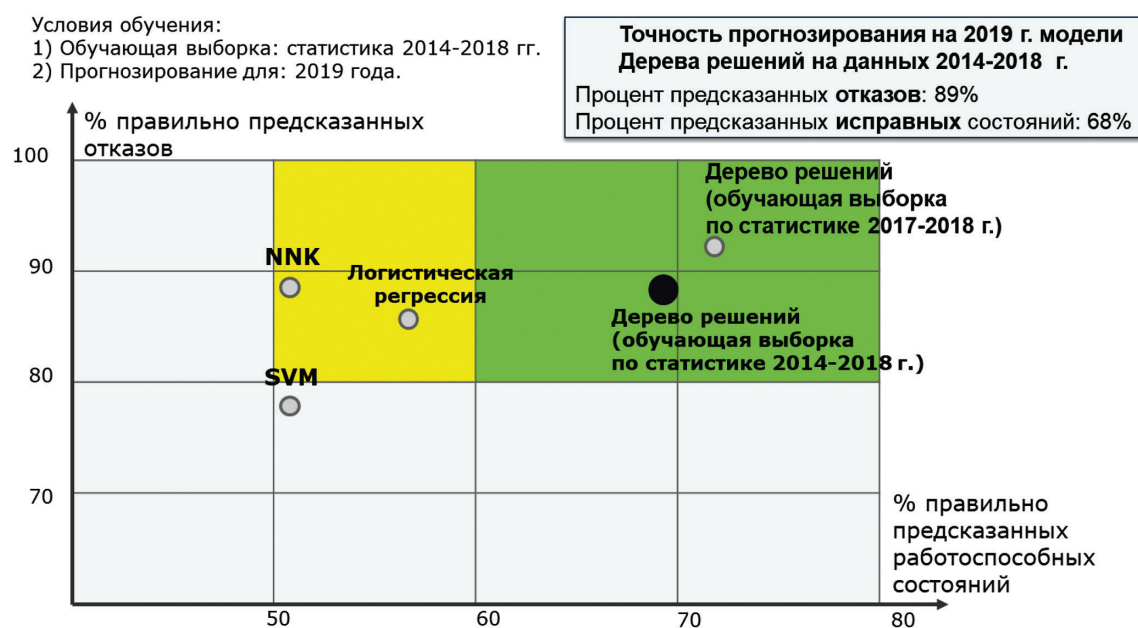


Рис. 5. Сравнение моделей по показателям качества



Табл. 3. Перечень объектов тестовой выборки в зоне недопустимого риска

Дата проверки	Дистанция пути	Эксплуатационный участок	Номер пути	Километр	Вероятность опасного отказа
29-JAN-19	9	2	1	979	0,55
29-JAN-19	9	1	1	969	0,51
14-JAN-19	9	2	1	979	0,48
14-JAN-19	9	1	1	969	0,48
29-JAN-19	9	2	1	1018	0,37
29-JAN-19	9	2	1	1003	0,28
14-JAN-19	9	2	1	1018	0,21
14-JAN-19	9	2	1	1003	0,17
23-JAN-19	20	1	1	36	0,003
25-JAN-19	20	2	1	36	0,0014

Результаты ранжирования моделей: ранг 1 – решающее дерево (обученное на выборке № 6), ранг 2 – случайный лес (обученный на выборке № 6).

В табл. 3 приведен перечень объектов верхнего строения пути, имеющих наибольшие вероятности опасного отказа (что соответствует наиболее высоким уровням риска) в январе 2019 г.

Анализ данных табл. 2 показывает, что для решения задачи классификации объектов наилучшие результаты могут быть получены в результате применения метода, основанного на решающих деревьях.

В последнем столбце табл. 3 приведена частота, с которой деревья решений классифицируют категорию объекта как «1», т.е. количество деревьев, определивших объект как «километр с опасным отказом ВСП» по отношению к общему числу построенных деревьев. По результатам работы алгоритма классификации для обучающей выборки необходимо выбрать порог значения вероятности отказа в зависимости от того, какая из ошибок классификации является приоритетной. Чем выше этот порог, тем реже объекты будут классифицироваться как «километр с опасным отказом» (величина ТР уменьшается, но увеличивается значение ТН). Чем порог ниже, тем меньше объектов «километр с опасным отказом» будет пропущено, но больше объектов с отсутствием опасного отказа («0») будет определено как имеющие опасный отказ («1») (возрастают значения ТР и FР). Для задачи классификации объектов ВСП важным является не пропустить возможный отказ. Пусть даже ценой увеличения количества объектов с отсутствием опасного отказа («0»), которые будут ложно определены как объекты с опасным отказом («1»).

По результатам классификации для обучающей выборки был выбран порог  $\bar{p}=0,15$ . На тестовой выборке это привело к тому, что из 10 объектов с опасным отказом 8 были классифицированы верно и 5 объектов без опасного отказа (с меткой «0») были также классифицированы как объекты с опасным отказом. Если бы

мы установили порог  $\bar{p}=0,10$ , то число правильно определенных объектов с опасным отказом («1») осталось бы неизменным, а число неверно классифицированных объектов без опасного отказа («0») увеличилось бы до 14. При  $\bar{p}=0,001$  все десять объектов с опасным отказом («1») были бы классифицированы верно, но при этом число неверно определенных объектов без опасного отказа («0») возросло бы до 251.

## 8. Заключение

В статье представлена методологическая основа прогнозирования редких опасных событий (отказов), которая может быть практически развита и реализована в ЕКП УРРАН ОАО «РЖД» в качестве составной системы. Эта система путем обработки большого массива входных данных должна в режиме реального времени прогнозировать неблагоприятные события на железнодорожном транспорте на некотором горизонте времени. Наполнение этой системы: комплекс математических моделей и методов, а также метрик для проверки качества построенных моделей.

На основе задачи прогнозирования отказов железнодорожного пути в зависимости от различного набора факторов отработана последовательность действий для формирования решения о целесообразности проведения дополнительных ремонтных работ на том или ином участке пути. Для этой цели проведено сравнение моделей с использованием предложенных метрик. Результатом категорирования объектов являлся вывод о наличии ключевых индикаторов и их значений раннего предупреждения факторов риска. Такой вывод основан на анализе соответствия фактических характеристик объекта и условий его эксплуатации случаям появления нежелательных событий и случаям отсутствия нежелательных событий. Предлагаемый набор методов и средств можно легко интегрировать с автоматизированной измерительной системой на базе железнодорожного подвижного состава.

## Библиографический список

1. Суконников Г.В. Применение технологии «интернет вещей» в ОАО «РЖД» [электронный ресурс] // Инновационный Дайджест: [сайт]. [2017]. URL: <http://www.rzd-expo.ru/innovation/novosti/1.pdf>
2. Бондаренко Ю.В., Куксо А.А., Маркевич И.Г. Информационные технологии в управлении системой диагностики железнодорожных путей // Сборник докладов IX международной научно-практической конференции студентов, аспирантов и молодых ученых: в 4 томах. 2018. С. 214–215.
3. Назаров Д.Г., Гура Д.А. О системах автоматизированного путеизмерительного контроля / Кубанский государственный технологический институт: Научные труды КУБГТУ. Краснодар, 2019. С. 135–146.
4. Куракина С.Г., Шмакова Е.Г. Автоматизация диагностики и мониторинга участников контактной сети на железных дорогах // Современные инновации. 2017. № 8(22). С. 15–17.
5. Замышляев А.М. Предпосылки для создания цифровой системы управления безопасностью движения // Надежность. 2019. № 4(71). С. 45–52.
6. Кузнецов Г.А., Крашенинников С.В., Крайсвитный В.П. и др. Модернизация системы ГИД «Урал-ВНИИЖТ» // Автоматика, связь, информатика. 2016. № 11. С. 15–19.
7. Замышляев А.М., Игнатов А.Н., Кибзун А.И. и др. Построение функциональной зависимости между рядом распределения количества вагонов в сходе и различными факторами движения // Надежность. 2018. Т. 18. № 1. С. 53–60.
8. Liu X., Saat M., Barkan C. Analysis of causes of major train derailment and their effect on accident rates // Transp. Res. Rec. J. Transp. Res. Board. 2012. Vol. 2289. P. 154–163.
9. Замышляев А.М., Игнатов А.Н., Кибзун А.И. и др. О нарушении безопасности движения, связанном с выходом в габарит соседнего пути подвижных единиц грузового поезда, сошедших с рельсов // Надежность. 2018. Т. 18. № 3. С. 39–45.
10. Lasisi A., Attoh-Okine N. Principal components analysis and track quality index: a machine learning approach // Transp. Res. Part C. Emerg. Technol. 2018. Vol. 91. P. 230–248.
11. Dr. Allan M. Zarembski Better railroading through Big Data [электронный ресурс] // Railway age: [сайт]. [2018]. URL: <https://www.railwayage.com/analytics/better-railroading-through-big-data/>
12. Thaduri A., Galar D., Kumar U. Railway assets: a potential domain for big data analytics // Proc. Comput. Sci. 2015. Vol. 53. P. 457–467.
13. Li Q., Zhong Z., Liang Z. et al. Rail inspection meets big data: methods and trends. // 18th International Conference on Network-Based Information Systems. 2015. P. 302–308.
14. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. М.: ДМК Пресс, 2015. 402 с.
15. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение М.: ДМК Пресс, 2018. 652 с.
16. Widodo A., Yang B.S. Support vector machine in machine condition monitoring and fault diagnosis // Mech. Syst. Signal Process. 2007. Vol. 21. P. 2560–2574.
17. Sun W., Chen J., Li J. Decision tree and PCA-based fault diagnosis of rotating machinery // Mech. Syst. Signal Process. 2007. Vol. 21. P. 1300–1317.
18. Cerrada M., Zurita G., Cabrera D. et al. Fault diagnosis in spur gears based on genetic algorithm and random forest // Mech. Syst. Signal Process. 2016. Vol. 70–71. P. 87–103.
19. Santur Y., Karakose M., Akin E. Random forest based diagnosis approach for rail fault inspection in railways // National Conference on Electrical, Electronics and Biomedical Engineering. 2016. P. 714–719.
20. Чистяков С.П. Случайные леса: обзор // Труды Карельского научного центра РАН. 2013. № 1. С. 117–136.
21. Hosmer D., Lemeshov S., Sturdivant R.X. Applied Logistic Regression. New York: John Wiley & Sons, 2013. 528 p.
22. Samworth R.J. Optimal weighted nearest neighbour classifiers // Ann. Statist. 2012. Vol. 40. № 5. P. 2733–2763.
23. Famurewa S.M., Zhang L., Asplund M. Maintenance analytics for railway infrastructure decision support // Journal Qual. Maint. Eng. 2017. № 23. P. 310–325.
24. Hu C., Liu X. Modeling Track Geometry Degradation Using Support Vector Machine Technique // 2016 Joint Rail Conference. 2016.
25. Бойко П.Ю., Быков Е.М., Соколов Е.И., Яроцкий Д.А. Применение машинного обучения к ранжированию инцидентов на Московской железной дороге // Информационные технологии и вычислительные системы. 2017. № 2. С. 43–53.
26. Jiang Y., Wang H., Tian G. et al. Fast classification for rail defect depths using a hybrid intelligent method // Optik (Stuttgart). 2019. Vol. 180. P. 455–468.
27. Gibert X., Patel V.M., Chellappa R. Deep multitask learning for railway track inspection // IEEE Trans. Intell. Transp. Syst. 2017. Vol. 18. P. 153–164.
28. Резницкий М.А., Аршинский Л.В. Программная реализация автоматизированной системы обнаружения дефектов верхнего строения пути на основе технологии свёрточных нейронных сетей // Молодая наука Сибири: электрон. науч. журн. 2018. № 1.
29. Lee J.S., Hwang S.H., Choi I.Y. et al. Prediction of track deterioration using maintenance data and machine learning schemes // J. Transp. Eng. Part A Syst. 2018. Vol. 144. P. 04018045-1:9.
30. Nakhaee M.C., Hiemstra D., Stoelinga M. et al. The Recent Applications of Machine Learning in Rail Track Maintenance: A Survey // Lecture Notes in Computer Science. 2019. P. 91–105.
31. Fawcett T. An introduction to ROC analysis // Pattern Recognition Letters. 2006. Vol. 27. P. 861–874.

## Сведения об авторах

**Игорь Б. Шубинский** – доктор технических наук, профессор, заместитель руководителя НТК АО «НИИАС», Москва, Российская Федерация, тел. +7 (495) 786-68-57, e-mail: igor-shubinsky@yandex.ru

**Алексей М. Замышляев** – доктор технических наук, заместитель Генерального директора АО «НИИАС», Москва, Российская Федерация, тел. +7 (495) 967-77-02, e-mail: A.Zamyshlaev@vniias.ru

**Ольга Б. Проневич** – начальник отдела АО «НИИАС», Москва, Российская Федерация, тел. +7 (985) 242-21-62, e-mail: oesune@rambler.ru

**Алексей Н. Игнатов** – кандидат физико-математических наук, старший преподаватель, Московский Авиационный институт, Москва, Российская Федерация, тел. +7 (906) 059-50-00, e-mail: alexei.ignatov1@gmail.com

**Евгений Н. Платонов** – кандидат физико-математических наук, доцент, Московский Авиационный институт, факультет «Прикладной математики и физики», Москва, Российская Федерация, тел. +7 (499) 158-45-60, e-mail: en.platonov@gmail.com

## Вклад авторов в статью

**Шубинский И.Б.** – формулировка требований к содержанию алгоритма подготовки данных для формирования выборок, целей каждого этапа. Формулировка требований к классификации методов машинного обучения на основании возможностей интерпретации результатов моделирования.

**Замышляев А.М.** – постановка цели работы, анализ проблемы и актуальности применения методов машинного обучения для прогнозирования появления опасных отказов верхнего строения пути, формирование выводов.

**Проневич О.Б.** – разработка алгоритма подготовки данных о состоянии железнодорожного пути для применения методов машинного обучения, алгоритма применения методов машинного обучения для прогноза появления опасных отказов железнодорожного пути.

**Игнатов А.Н.** – предварительная обработка и анализ данных для расчета.

**Платонов Е.Н.** – обзор методов машинного обучения и их применение в анализе дефектов железнодорожного пути. Постановка задачи классификации.

Работа выполнена при поддержке РФФИ (проект № 20-07-00046А).