

На правах рукописи



КУЛАГИН МАКСИМ АЛЕКСЕЕВИЧ

**ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АНАЛИЗА И
ПРОГНОЗИРОВАНИЯ НАРУШЕНИЙ ПРИ УПРАВЛЕНИИ
ПОДВИЖНЫМ СОСТАВОМ**

2.9.8. Интеллектуальные транспортные системы

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата технических наук

Москва – 2022

Работа выполнена в федеральном государственном автономном образовательном учреждении высшего образования «Российский университет транспорта».

Научный руководитель: доктор технических наук, профессор
Сидоренко Валентина Геннадьевна

Официальные оппоненты: **Ковшов Евгений Евгеньевич**,
доктор технических наук, профессор,
акционерное общество «Научно-исследовательский и конструкторский институт монтажной технологии – Атомстрой», научно-инженерная и образовательная лаборатория цифровых компьютерных систем и автоматизации, начальник лаборатории

Хромов Игорь Юрьевич,
кандидат технических наук, общество с ограниченной ответственностью «Программный инжиниринг и технологии», центр разработки и сопровождения корпоративных систем управления сервисом подвижного состава, старший бизнес-аналитик

Ведущая организация: Федеральное государственное бюджетное образовательное учреждение высшего образования «Петербургский государственный университет путей сообщения Императора Александра I»,
г. Санкт-Петербург

Защита диссертации состоится 21 сентября 2022 г. в 11:00 часов на заседании диссертационного совета 40.2.002.05 на базе федерального государственного автономного образовательного учреждения высшего образования «Российский университет транспорта» по адресу: 127994, ГСП–4, г. Москва, ул. Образцова, 9, стр. 9, ауд. 2505.

С диссертацией можно ознакомиться в научной библиотеке и на сайте РУТ (МИИТ) www.miit.ru.

Автореферат разослан «__» июня 2022 года.

Ученый секретарь

диссертационного совета



Журавлев Илья Александрович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. «Человеческий фактор» имеет достаточно высокое влияние на безопасность в транспортной отрасли. Оценка надежности человека всегда была важной проблемой для научного сообщества, экспертов по безопасности и инженеров. Большинство методов оценки надежности человека субъективны, а данные о человеческом факторе неточны. С ростом объема информации о поведении человека и появлением способов её обработки возникла возможность обеспечить объективный подход к оценке надежности человека, основанный на статистической оценке данных о нём. В области компьютерных технологий высоко зарекомендовали себя методы машинного обучения, которые используются в различных областях науки для эффективного анализа большого объема данных, прогноза событий, детектирования и распознавания образов, распознавания речи.

В ОАО «РЖД» на данный момент при расследовании причин, повлиявших на возникновение грубых транспортных происшествий, часто устанавливается повторяемость нарушений при управлении подвижным составом и другие факторы, которые явились предпосылками к происшествию. Из-за отсутствия системы прогнозирования эти предпосылки выявляются апостериорно. Синтез способов определения уровня надежности машиниста и прогнозирования совершения нарушения при управлении подвижным составом является актуальной задачей. Выявление предпосылок совершения грубых нарушений при управлении подвижным составом позволит компании значительно повысить уровень безопасности движения и уменьшить возможные финансовые издержки. Внедрение математических моделей оценки деятельности машиниста при управлении подвижным составом в автоматизированных системах управления, используемых в ОАО «РЖД», поможет руководителям различных уровней принимать своевременные решения и формировать профилактические мероприятия для снижения вероятности возникновения нарушений.

Степень разработанности темы исследования. В мире накоплен значительный опыт в области управления безопасностью и анализа влияния «человеческого фактора» на безопасность движения поездов.

Большой вклад в управление движением поездов и обеспечение его безопасности в ОАО «РЖД», а также анализ «человеческого фактора»,

влияющего на технологический процесс, внесли ученые России. Авторами работ, связанных с алгоритмами и методами управления безопасностью в ОАО «РЖД», являются Баранов Л.А., Бестемьянов П.Ф., Гапанович В.А., Горелик А.В., Замышляев А.М., Лисенков В.М., Розенберг Е.Н., Ройзнер А.Г., Сапожников В.В., Шалягин Д.В., Шаманов В.И., Шихер Я.Г., Шубинский И.Б., Ярковский Ф.В. и др. Вопросами оценки влияния «человеческого фактора» на технологический процесс занимались в первую очередь Аксёнов В.А., Губинский А.И., Дружинин Г.В., Завьялов А.М., Захаревич А.С., Катцын Д.В., Котик М.А., Маринов М.Л., Никифорова Н.Б., Падерно П.И., Ротштейн А.П., Салвенди Г., Синякина И.Н. и др. Вопросами разработки систем контроля состояния человека, оказывающих влияние на технологический процесс, занимались Емельянов А.М., Захаревич А.С., Катцын Д.В., Киселева И.А., Котик М.А., Подопрыголова О.Н., Семенистая Е.С., Симонович Н.Е., Шахнарович В.М. и др.

В работах зарубежных авторов проводится общий анализ практических результатов применения разнообразного математического аппарата для анализа и оценки «человеческого фактора». Примером могут служить работы *Blackman H., Cooper S.E., Corlett E.N., Forester J.A., Galyean W.J., Gertman D., Guttmann H.E., Hollnagel E., Kirwan B., Li P.C., Ramey-Smith A.M., Shirali G.A., Swain A.D., Wiegmann D.A., Wilson J.R., Wreathall J.A.* и др.

Объектом исследования являются факторы, оказывающие влияние на возникновение нарушений при управлении подвижным составом, и мероприятия по обеспечению безопасности управления движением поездов на железнодорожном транспорте.

Предмет исследования. Интеллектуальная система анализа и прогнозирования нарушений при управлении подвижным составом.

Целью диссертации является создание математического, информационного и программного обеспечения интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом, реализующей формирование профилактических мероприятий, направленных на минимизацию числа этих нарушений.

Для достижения поставленной цели необходимо решить следующие **задачи:**

1. Анализ и классификация существующих методов и систем оценки надежности человека и влияния «человеческого фактора» на безопасность транспортных систем.

2. Разработка принципов построения и математических моделей интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом.

3. Формирование множества показателей работы машиниста, используемых интеллектуальной системой анализа и прогнозирования нарушений при управлении подвижным составом, определение их влияния на безопасность движения поездов, выбор алгоритмов обработки данных и определение критериев оценки качества работы машиниста.

4. Разработка алгоритма формирования групп надежности машинистов и перечня профилактических мероприятий, направленных на повышение надежности машинистов.

5. Реализация и анализ результатов применения интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом в рамках автоматизированной системы управления, используемой в ОАО «РЖД».

Научная новизна состоит в том, что:

– Выполнена классификация методов и систем оценки надежности человека и влияния «человеческого фактора» на безопасность транспортных систем, позволившая определить методы искусственного интеллекта как перспективные для решения задач прогнозного анализа работы машиниста.

– Построены и апробированы на практике математические модели, отличающиеся от ранее известных моделей анализа и оценки влияния «человеческого фактора» на технологический процесс применением ансамбля базовых алгоритмов машинного обучения, интегрирующего решение нескольких взаимосвязанных задач, а именно расчёта вероятности совершения нарушения в ближайшем будущем и формирования списка прогнозируемых нарушений, и адаптацией к специфике решаемой задачи.

– Для задач, решаемых интеллектуальной системой анализа и прогнозирования нарушений при управлении подвижным составом, построены математические модели машинного обучения, дающие возможность определить уровень надежности машиниста, вероятность

совершения нарушения и сформировать перечень рекомендуемых для машиниста профилактических мероприятий.

– Сформировано расширенное множество показателей, характеризующих работу, отдых и физическое состояние машиниста, учитывающих динамику его поведения и наличие статистических связей между этими показателями, позволяющих определить влияние надежности машиниста на безопасность движения поездов и организовать работу интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом.

– Показано влияние отдельных показателей, характеризующих работу машиниста, на нарушения, совершаемые при управлении подвижным составом.

– Разработан алгоритм формирования перечня рекомендуемых машинисту профилактических мероприятий с учётом результатов работы математических моделей интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом, отличающийся от существующих снижением влияния «человеческого фактора» на оценку уровня надежности машиниста.

– Разработаны и реализованы на практике требования к архитектуре интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом с учетом необходимости работы с большими данными и обеспечения простоты сопровождения, разработки и развертывания.

Теоретическая и практическая значимость состоит в том, что:

– Выполнен анализ методов и систем оценки надежности человека и влияния «человеческого фактора» на безопасность транспортных систем, дающий возможность сформулировать требования к структуре и функционалу интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом и определить множество методов решения задач, поставленных перед системой.

– Построены математические модели анализа исторических данных и оценки работы машиниста, позволяющие в отличие от классических методов экспертной оценки производить более объективный и обобщенный прогнозный анализ поведения машиниста в будущем и формировать базис для

определения перечня рекомендуемых машинисту профилактических мероприятий.

– Сформировано расширенное множество показателей работы машиниста, используемых в качестве исходных данных при построении математических моделей машинного обучения и статистического анализа, применяемых для решения задач интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом в рамках разработанной структуры.

– Разработан метод оценки и отбора показателей работы машинистов для создания моделей машинного обучения с целью решения задачи прогнозирования нарушений машинистом и расчета уровня надежности с учетом использования технологий по обработке большого объема данных (*BigData*).

– Разработана математическая модель формирования перечня рекомендуемых машинисту профилактических мероприятий, основанная на использовании результатов работы алгоритмов по расчету вероятности совершения грубого нарушения безопасности и формированию списка вероятных нарушений, которая при её внедрении на программном уровне позволит автоматизировать этот процесс.

– Сформулированы и реализованы на практике требования использования и внедрения математических моделей, способа формирования групп надежности и перечня рекомендуемых машинисту профилактических мероприятий для автоматизированной системы управления, внедренной в ОАО «РЖД», что позволило создать интеллектуальную систему анализа и прогнозирования нарушений при управлении подвижным составом, снижающую влияние «человеческого фактора».

– Построена архитектура и разработаны требования к разрабатываемым приложениям и их интеграции с базами данных для интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом машинистом, основывающиеся на комбинации нескольких подходов, используемых в области *BigData*, а именно сочетания архитектур *IoT* и микросервисной.

Методология и методы исследований. Результаты диссертационной работы получены на основе использования системного анализа, теории

управления, теории систем искусственного интеллекта, теории машинного обучения, теории вероятности, теории баз данных и разработки программного обеспечения.

Положения, выносимые на защиту:

– Классификация методов и систем оценки надежности человека и влияния «человеческого фактора» на безопасность транспортных систем.

– Математические модели решения задач интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом, основанные на использовании исторических данных и комбинации моделей машинного обучения по расчету вероятности совершения нарушения, построению списка вероятных нарушений и формированию перечня рекомендуемых машинисту профилактических мероприятий.

– Состав множества показателей работы машиниста, используемых в интеллектуальной системе анализа и прогнозирования нарушений при управлении подвижным составом, учитывающих динамику его поведения и наличие статистических связей между этими показателями.

– Алгоритм формирования перечня профилактических мероприятий, рекомендуемых машинисту, основанный на использовании результатов работы комбинации моделей по расчету вероятности факта совершения нарушения и формированию списка вероятных нарушений.

– Требования к построению и внедрению разработанных моделей, алгоритмов и способов оценки деятельности машиниста в автоматизированные системы управления, используемые в ОАО «РЖД».

– Микросервисная архитектура программного обеспечения интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом, которая позволила организовать эффективную работу системы с точки зрения процессов хранения, получения и обработки данных, а также обеспечить простоту сопровождения, разработки и развертывания.

Степень достоверности полученных результатов подтверждается результатами проверки адекватности использованных моделей и обусловлена корректностью постановок задач, обоснованностью принятых допущений, высоким качеством работы математических моделей на основе данных, полученных из различных автоматизированных систем управления,

внедренных в ОАО «РЖД», совпадением результатов прогнозирования с фактическими нарушениями в управлении движением поездов.

Апробация результатов работы. Основные результаты работы докладывались на заседаниях кафедры «Управление и защита информации» («УиЗИ») (РУТ(МИИТ) 2017 - 2021 гг.), на двух научно-практических конференциях «Неделя Науки» (РУТ(МИИТ), 2018 - 2019 гг.), на двух научно-практических конференциях «Безопасность Движения Поездов» (РУТ(МИИТ), 2017 - 2018 гг.), на трех международных конференциях «Проблемы управления безопасностью сложных систем» (ИПУ РАН, 2018, 2019, 2021 г.), на международной конференции «*Intelligent Information Technologies for Industry*» (Russia (Sochi), 2018 г.), на одной международной конференции «*East-West Design & Test Symposium*» (Varna, Bulgaria, 2020 г.).

Структура и объём работы. Диссертация состоит из введения, 4 глав, заключения, 8 приложений, списка литературы, включающего 167 наименований. Материалы диссертации содержат 159 страниц основного текста, 67 страниц приложений, 52 рисунка, 14 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Во введении показана актуальность работы, определены цель и задачи исследования.

В первой главе проведен обзор существующих методов и систем оценки деятельности работников транспортных служб. Приведены основные термины и определения используемых алгоритмов и методов машинного обучения. Выявлено, что в большинстве рассмотренных методов производится расчет вероятности возникновения неблагоприятного события классическими аналитическими методами, в частности методами экспертных оценок. Отмечено, что оценку человеческой деятельности необходимо проводить с использованием нелинейных математических моделей и полученных исторических данных. Показана актуальность проводимых исследований, сформулированы основные направления исследования и определена необходимость комплексного решения поставленных задач. Для достижения поставленных цели и задач построена структурная схема интеллектуальной системы анализа и прогнозирования нарушений при управлении подвижным составом (ИСАПНУПС) (Рисунок 1).

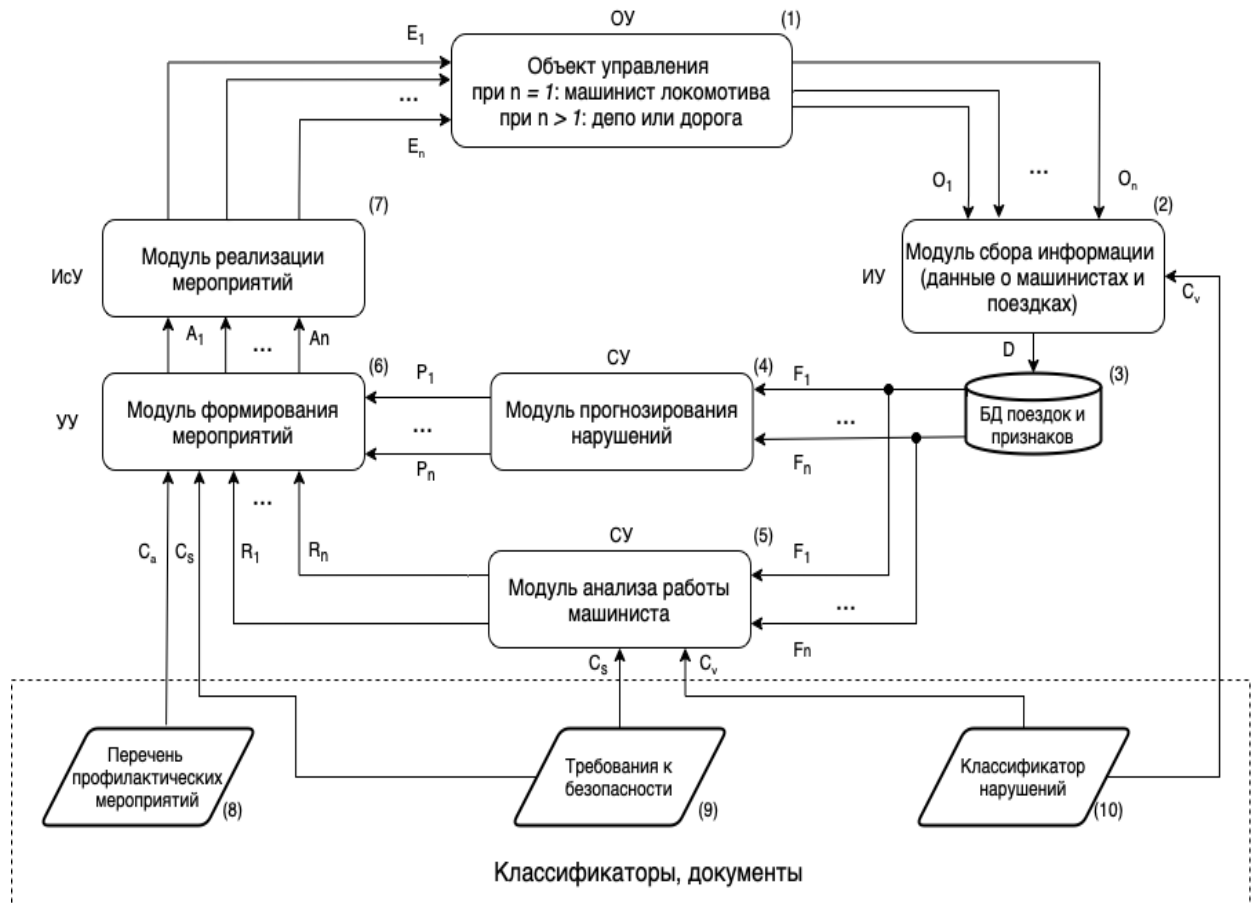


Рисунок 1 – Структурная схема ИСАПНУПС и формирования списка рекомендуемых мероприятий, где ОУ – объект управления; ИУ – измерительное устройство; СУ – сравнивающее устройство; УУ – устройство управления; ИсУ – исполнительное устройство

На вход объект управления поступают управляющие воздействия $[E_1, E_2, \dots, E_n]$, а на выходе формируется массив реакций, которые совершает объект управления – $[O_1, O_2, \dots, O_n]$. Модуль сбора информации формирует массив поездок машиниста D . На выходе базы данных формируется вектор признаков и характеристик машинистов – $[F_1, F_2, \dots, F_n]$. Модуль прогнозирования нарушений рассчитывает вектор вероятностей $[P_1, P_2, \dots, P_n]$ совершения нарушений машинистами. На выходе модуля анализа работы машиниста формируется вектор значений групп риска для каждого машиниста $[R_1, R_2, \dots, R_n]$ и критерии для формирования перечня профилактических мероприятий. Модуль формирования мероприятий создает перечень мероприятий и действий, направленных на повышение уровня безопасности движения поездов $[A_1, A_2, \dots, A_n]$.

Дано определение понятия надежность машиниста, как способность не совершать в течение R поездок нарушений при управлении подвижным

составом, где $R \in [1 \dots N]$. Вероятность отсутствия нарушений при управлении подвижным составом может быть рассчитана через вероятность обратного события:

$$P(R) = 1 - P_N(R), \quad (1)$$

где $P_N(R)$ – вероятность нарушения при управлении подвижным составом в течение R поездок. Определение этой величины требует знаний о корреляционной зависимости между показателями машиниста и совершаемыми нарушениями, анализ которой представлен в диссертации.

Вторая глава посвящена анализу работы машиниста и его обязанностей, сбору и анализу данных, используемых в исследовании, преобразованию категориальных признаков, масштабированию и нормализации числовых признаков, поиску линейно- и нелинейно-зависимых признаков. В качестве основных источников информации о работе локомотивных бригад использованы данные из 7 автоматизированных систем управления, внедренных в ОАО «РЖД». Определен перечень используемых ИСАПНУПС источников информации и разработан алгоритм сбора и объединения данных по машинистам. В исследовании использованы данные 4 737 561 поездок по 16 дорогам, 143 депо для 56 605 машинистов и 33 575 помощников машинистов. Разработан алгоритм расчёта признаков, характеризующих динамику изменения качества работы машиниста и рассчитываемых по данным предыдущих поездок и текущим характеристикам машиниста.

Например, на основе данных по нарушениям, выявленным при расшифровке скоростемерных лент, рассчитаны тенденции изменения значений показателей за 7–30, 30–90, 90–180 дней:

$$d = \frac{\Delta n}{\Delta t} = \frac{n_2 - n_1}{t_2 - t_1}, \quad (2)$$

где Δn – разница между значениями показателей N_{grub} , N_{negrub} , N_{povt} , N_{ves} за разные промежутки времени Δt ; Δt – промежутки времени, за которые рассчитываются показатели; n_2 – параметр, который может принимать значения N_{grub} , N_{negrub} , N_{povt} , N_{ves} за 180, 90, 30 дней; n_1 – параметр, который может принимать значения N_{grub} , N_{negrub} , N_{povt} , N_{ves} за 90, 30, 7 дней; N_{grub} , N_{negrub} , N_{povt} – количество грубых, негрубых и повторяющихся нарушений за 7, 30, 90, 180 дней от текущей поездки; N_{ves} – количество баллов по нарушениям за 7, 30, 90, 180 дней от текущей поездки.

При анализе качества расследования установлено (Рисунок 2), что в группе грубых нарушений, в которых вина машиниста не может быть не

установлена, вина подтверждается в среднем по Дирекции тяги в 50% случаев. Наиболее качественно расследуют причины нарушений с подтверждением вины на двух дорогах, где отношение грубых нарушений с виной к общему количеству грубых нарушений составляет 0.87–0.9. Лучшие показатели обведены эллипсами.

С использованием методов экспертных оценок определены значимости нарушений, которые используются в качестве признаков и меток в процессе обучения алгоритмов машинного обучения. Веса определялись с использованием коэффициента конкордации Кендалла. Полученные результаты показывают, что гипотезу о том, что мнения экспертов не согласованы, можно отвергнуть на уровне значимости $\alpha = 0.05$. В итоге можно сделать вывод, что оценка значимостей нарушений на основе принимаемых мер является более объективной в сравнении с оценкой нарушений напрямую.

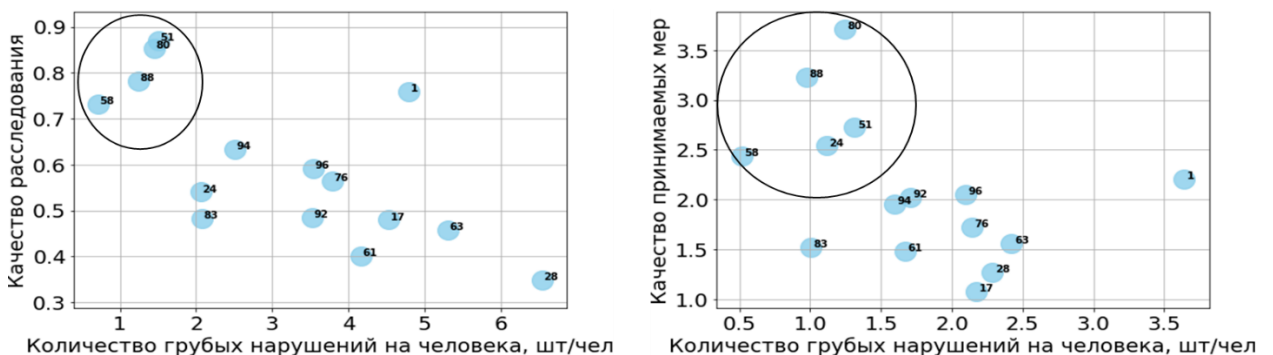


Рисунок 2 – Зависимость количества грубых нарушений от качества расследования нарушений (отношение грубых нарушений с виной к общему количеству грубых нарушений) и качества принимаемых мер по дорогам (среднее значение веса за мероприятия, принятые по результатам расследования)

Произведена оценка влияния признаков, используемых в исследовании, на целевую метку и друг на друга с использованием статистических методов проверки гипотез, адаптированных под большой объем данных (Таблица 1). В качестве статистических критериев использовались: t -критерий Стьюдента и критерий Манна-Уитни.

Таблица 1 – Статистическая оценка влияния показателей на появление грубых нарушений

Группа признаков	Значение p -value	Вывод при уровне значимости 0.05
<i>t</i> -критерий Стьюдента: нулевая гипотеза – равенство средних значений ($H_0: \mu_1 = \mu_2$), альтернативная гипотеза – неравенство средних значений ($H_1: \mu_1 \neq \mu_2$) в выборках с нарушениями и без нарушений		
Время работы за неделю/месяц до нарушения	0.000246 / 0.00334	Гипотезу H_0 можно отклонить в пользу альтернативы, т. е. среднее время работы у машинистов с нарушениями и без нарушений – отличается.
Количество баллов за последнюю плановую проверку знаний	0.0494	Гипотезу H_0 можно отклонить в пользу альтернативы, т. е. среднее количество баллов, набранных за плановую проверку знаний влияет на появление нарушений.
Количество уникальных помощников за последние 7/30 дней	0.0028 / 0.0014	Гипотезу H_0 можно отклонить в пользу альтернативы, т. е. среднее количество уникальных помощников у машиниста влияет на нарушения.
Критерий Манна-Уитни: нулевая гипотеза – распределение совпадают ($H_0: F_{X_1}(x) = F_{X_2}(x)$), альтернативная гипотеза – между распределениями существует сдвиг ($H_1: F_{X_1}(x) = F_{X_2}(x + \Delta), \Delta \neq 0$) в выборках с нарушениями и без нарушений		
Количество симптомов дизадаптации/срывов адаптации	0.0272 / 0.0168	Гипотезу H_0 можно отклонить в пользу альтернативы, т. е. физическое состояние машиниста, оцененное по медицинским характеристикам, влияет на появление нарушений.
Количество случаев расхода топливно-энергетических ресурсов (ТЭР) по отношению к нормальному на 20% за 7/30 дней	0.0052 / 0.0021	Гипотезу H_0 можно отклонить в пользу альтернативы, т. е. среднее количество случаев перерасхода топлива на 20% – отличается.
Средний балл по неплановым проверкам знаний за последние 30 дней	0.0023	Гипотезу H_0 можно отклонить в пользу альтернативы, т. е. среднее количество баллов, набранных за неплановые проверки знаний влияет на появление нарушений.

Влияние количества грубых нарушений набранных за последние 7/30/90 дней	$1.41 * 10^{-37}$ / $8.34 * 10^{-39}$ / $1.56 * 10^{-39}$	Гипотезу H_0 можно отклонить в пользу альтернативы, т. е. среднее количество нарушений, набранных в прошлом, влияет на нарушения.
Частота смены норм-участка за последние 30/90 дней	0.07763 / 0.09114	Гипотезу H_0 нельзя отклонить в пользу альтернативы, т. е. можно заключить об отсутствии теоретической значимости частоты смены норм-участка машиниста за 30 и 90 дней.

Кроме статистической оценки требуется рассмотреть зависимость показателей от среднего количества грубых нарушений. В автореферате приведен один пример, остальные представлены в тексте диссертации (Рисунок 3). Из полученных графиков можно заключить, что есть зависимость между средним временем работы машиниста и количеством грубых нарушений. Значит данный показатель можно использовать для прогнозирования нарушений.

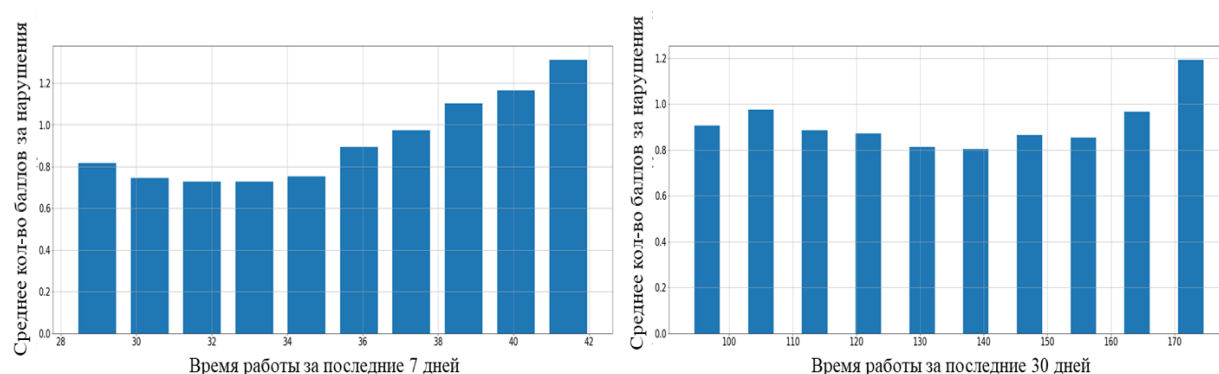


Рисунок 3 – Зависимость средних значений количества грубых нарушений от времени работы за последние 7 и 30 дней

Составлена матрица корреляции между показателями, которая позволила выявить линейную зависимость и исключить из обучающей выборки линейно-зависимые показатели. Это позволило наиболее эффективно подобрать матрицу признаков.

В третьей главе разработаны модель прогнозирования нарушений при управлении подвижным составом в ближайшем будущем, алгоритм

формирования рейтинга машинистов и модель прогнозирования конкретных нарушения при управлении подвижным составом в ближайшем будущем.

Модель прогнозирования нарушений в работе машинистов решает задачу бинарной классификации по прогнозированию совершения нарушения машинистом в ближайшие K дней. В качестве тестовой выборки использовалось 846 тыс. поездок. В качестве меток для задачи прогнозирования нарушений при управлении подвижным составом были выбраны грубые нарушения с установленной виной машиниста, которые могут быть совершены в ближайшие 3, 7, 10, 20, 30, 60 дней. Выборка меток машинистов является несбалансированной, что учтено в алгоритме. Наилучшие результаты из рассмотренных вариантов показала комбинация алгоритма градиентного бустинга над решающими деревьями и способа преобразования категориальных признаков на основе частотного представления и масштабирования. Составлена соответствующая таблица сопряжённости (Рисунок 4).

		Прогноз			
		Без нарушений	С нарушениями		
Факт	Без нарушений	0.97	0.03	Без нарушений	
	С нарушениями	0.24	0.76		

а)

		Прогноз			
		Без нарушений	С нарушениями		
Факт	Без нарушений	0.99	0.01	Без нарушений	
	С нарушениями	0.18	0.82		

б)

Рисунок 4 – Матрица ошибок для горизонтов прогнозирования: а) 30 дней; б) 60 дней

В качестве дополнительных результатов с помощью модели градиентного бустинга определены наиболее значимые показатели работы машиниста. Среди них можно выделить: последнее совершенное нарушение, вес, набранный за нарушения за последние 30 и 180 дней, уровень знаний и др.

Уровень значимости признака imp_F рассчитывается исходя из его влияния на изменение значения прогноза, т. е. насколько в среднем изменяется прогноз при изменении значения признака. Расчет выполняется на основе пороговой функции и значения конкретного показателя F . Чем больше значимость признака, тем в среднем больше он влияет на изменение значения прогноза. При реализации градиентного бустинга над решающими деревьями строится N решающих деревьев. При прохождении вектора признаков

машиниста через решающие деревья от корня к листьям в каждом узле проверяются условия разделения. Если условие выполнено (это условие зависит от значения F), объект переходит в левое поддерево; в противном случае он переходит в правое:

$$imp_F = \sum_{ts, ls_F} \left(k_1 - \frac{k_1 \cdot r_1 + k_2 \cdot r_2}{r_1 + r_2} \right) \cdot r_1 + \left(k_2 - \frac{k_1 \cdot r_1 + k_2 \cdot r_2}{r_1 + r_2} \right) \cdot r_2, \quad (3)$$

где r_1, r_2 – общий вес объектов в левом и правом листе соответственно. Этот вес равен количеству объектов в каждом поддереве; k_1, k_2 – значение прогноза в левом и правом поддереве соответственно; ts – количество решающих деревьев; ls_F – количество листьев, которые участвовали в разделении признака F .

Разработана математическая модель прогнозирования нарушений при управлении подвижным составом на основе информации о ранее совершенных нарушениях и признаках, характеризующих деятельность машиниста. Работа нейронной сети апробирована на тестовой выборке размером 5-10 тысяч машинистов. В качестве метрик оценок качества использовалась точность попадания нарушения в топ 10 и 20 нарушений в спрогнозированном векторе нарушений $Accuracy@topK$. В результате была обучена нейронная сеть, в которой в качестве входного слоя выступает вектор признаков, сформированный из трёх различных наборов данных (U – вектор, характеризующий машиниста, V – вектор, характеризующий нарушения, и прочие признаки машиниста, указанные в Приложении 1 диссертации). У данной модели есть ряд ограничений, а именно, необходимо, чтобы машинист совершил какое-либо нарушение, тогда появляется возможность рассчитать вектор прогнозируемых нарушений.

Разработан алгоритм формирования рейтинга машиниста путем анализа отклонения показателей машиниста от показателей «эталонного» машиниста. Под «эталонным» машинистом подразумевается вектор признаков, которыми должен обладать каждый машинист согласно установленным в ОАО «РЖД» правилам. Использовались следующие метрики для построения рейтинга: коэффициент корреляции, метрика Минковского, косинусное расстояние, расстояние Махалонобиса. Рейтинг используется для определения перечня рекомендуемых профилактических мероприятий, так как в данный подход позволяет определить, какой показатель внёс наибольший вклад в отклонение вектора показателей машиниста от целевого вектора.

В рамках диссертации разработан алгоритм объединения трёх математических моделей и формирования профилактических мероприятий – алгоритм, который последовательно производит вычисления для каждой модели и передаёт результат для вычисления надёжности машиниста (Рисунок 5). Если за N дней нарушений не было (условие в Блоке 6 не выполняется), то запускается процесс прогнозирования нарушения в ближайшее время (Блок 7). Прогноз совершения грубого нарушения сравнивается с допустимым порогом r . Если P не превысило порог r , то машинист допускается до поездки. Если за N дней нарушения были, то запускается рекомендательная подсистема (Блок 8), которая формирует список нарушений, которые может совершить данный машинист в будущем.

Расчёт P производится с использованием формулы полной вероятности путём перехода к расчёту вероятности обратного события:

$$P = 1 - \prod_{i=1}^K (1 - p_i) \quad (4)$$

где K – длина вектора нарушений, который формируется в математической модели расчёта вероятности; p_i – вероятность совершения нарушения в ближайшие 1, 7, 10 и 30 дней. Интерпретировать P можно как уровень надёжности машиниста.

Разработан способ формирования групп надёжности машинистов на основе квантилей (0.9, 0.7, 0.5) распределения оценок вероятности совершения нарушения и определения уровня надёжности в ближайшем будущем. В итоге было сформировано 4 группы надёжности: высокий уровень надёжности (0.96; 1]; допустимый уровень надёжности (0.73; 0.96]; недопустимый уровень надёжности (0.39; 0.73]; критический уровень надёжности (0; 0.39].

В четвертой главе рассмотрены способы формирования мероприятий на основе поиска сходства между показателями «эталонного» и исследуемого машинистов, а также результаты прогноза нарушений. Каждое нарушение характеризуется двумя группами факторов: общая характеристика (грубое или негрубое нарушение, с нарушением регламента, с возможностью нарушения безопасности, с нарушением безопасности) и человеческий фактор (низкий уровень знаний, недостаток опыта, невнимательность, отвлечение, спешка, халатность).

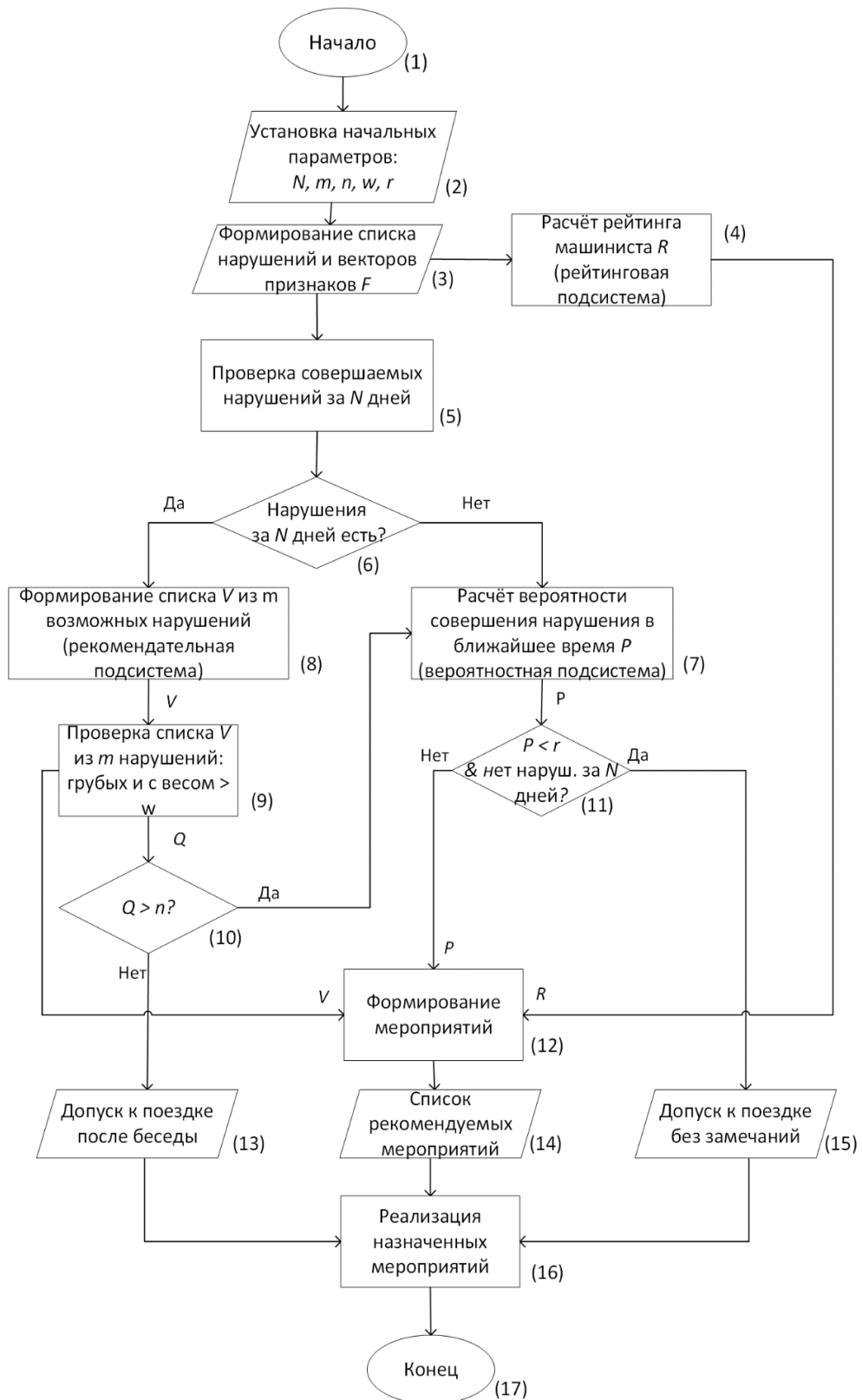


Рисунок 5 – Схема алгоритма объединения моделей и формирования рекомендуемых профилактических мероприятий

Пример формирования мероприятий для двух реальных машинистов с уникальными идентификаторами 5 и 10 в рамках функционирования разработанных алгоритмов в рамках автоматизированной системы

«Доверенная среда локомотивного комплекса» представлен ниже (Рисунок 6). В рассматриваемых примерах при формировании списка рекомендуемых профилактических мероприятий учитываются вероятность совершения нарушения в ближайшие 30 дней и фактические показатели.

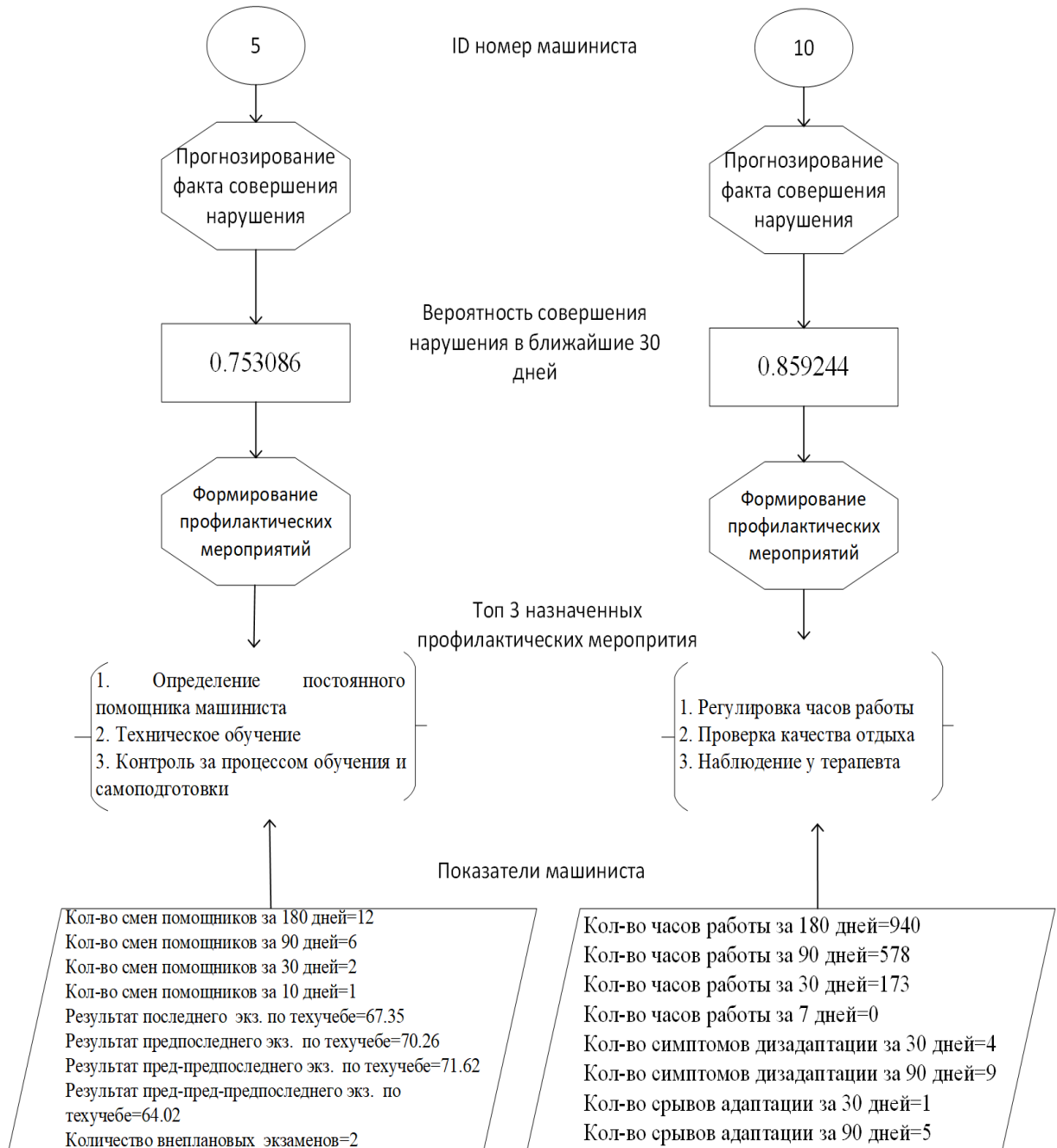


Рисунок 6 – Результаты работы ИС на примере двух машинистов

Разработана архитектура и требования к построению гетерогенной системы сбора и обработки большого массива информации, поступающей из различных источников, которой является ИСАПНУПС.

ЗАКЛЮЧЕНИЕ

1. Выполнена классификация методов и систем оценки надежности человека и влияния «человеческого фактора» на безопасность транспортных систем, позволившая выбрать перспективные направления и методы решения задач прогнозного анализа работы машиниста, к которым относятся методы искусственного интеллекта. В рамках ИСАПНУПС применение этих методов позволяет повысить точность прогноза, уменьшить влияние «человеческого фактора» на безопасность движения поездов, повысить эффективность управления, парирующего влияние «человеческого фактора» на безопасность движения.

2. Определено множество алгоритмов машинного обучения, перспективных для определения уровня надежности машиниста и прогноза возникновения грубого нарушения.

3. Разработан алгоритм объединения математических моделей, используемых в ИСАПНУПС для оценки влияния «человеческого фактора» на технологический процесс с точки зрения прогнозирования нарушения в ближайшем будущем, формирования списка прогнозируемых нарушений и расчета рейтинга машиниста.

4. Произведен анализ результатов на 846 тыс. поездках, который позволил выявить, что модель прогнозирования грубого нарушения на ближайшие 30 дней, основанная на решающих деревьях, показывает наилучшие результаты при решении задачи классификации. За май-июнь 2021 года было верно классифицировано 77% (2270) случаев факта совершения нарушения, 23% (687) случаев пропущено, то есть тех случаев, в которых машинист мог совершать нарушение, но при этом у него был низкий уровень надежности.

5. Разработан алгоритм расчета вероятности появления конкретного нарушения, который базируется на использовании гипотезы о том, что появление нарушений влечёт за собой появление подобных. Точность работы модели составила 62% при попадании прогнозируемого нарушения в множество 20-и наиболее вероятных нарушений. Данная модель участвует в алгоритме формирования набора точечных рекомендуемых машинисту профилактических мероприятий.

6. Построен алгоритм формирования перечня рекомендуемых профилактических мероприятий, позволяющий на базе результатов работы моделей и показателей, характеризующих работу машиниста, сформировать

индивидуальный набор мероприятий, направленных на уменьшение вероятности возникновения нарушения. Способ формирования мероприятий заключался в комбинации нескольких подходов: поиска сходства между «эталонным» машинистом и фактическими показателями исследуемого машиниста на основе информации о прогнозируемых и совершенных нарушениях при управлении подвижным составом.

7. Выделено две основные группы мероприятий: «краткосрочные» и «долгосрочные», что позволило в ИСАПНУПС сформировать меры, рекомендуемые машинисту перед и после поездки.

8. Разработаны требования, необходимые для внедрения математических моделей, алгоритмов формирования групп надежности и перечня рекомендуемых машинисту профилактических мероприятий, что позволило создать ИСАПНУПС.

9. Разработана архитектура программного обеспечения взаимодействия приложений с базами данных на основе комбинации нескольких подходов, используемых в области *BigData*, учитывающая сочетание архитектур *IoT* и микросервисной. Данная комбинация позволяет выделить явным образом в системе бизнес-процессы и сформировать потоковую обработку данных на ежедневной основе.

10. В качестве перспектив развития представленного диссертационного исследования определены следующие направления: разработка и верификация экономического критерия оценки эффективности используемых моделей и рекомендуемых профилактических мероприятий в рамках компании ОАО «РЖД»; повышение качества работы моделей, в частности, путем поиска новых значимых показателей, внедрения новых алгоритмов обработки данных и применения математического аппарата управления функциональной надежностью; повышение уровня автоматизации процесса расшифровки скоростемерных лент и учёта данных с регистраторов переговоров машиниста и помощника машиниста, машиниста и диспетчера.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

а) в рецензируемых научных изданиях:

1. Кулагин, М. А. Квалификация машинистов как фактор повышения надежности работы электроподвижного состава / М.А. Кулагин, В.Г. Сидоренко // Наука и техника транспорта. – 2018. – № 4. – С. 70-76.

2. Кулагин, М. А. Автоматизация сопровождения пользователей информационных и автоматизированных систем на базе искусственного интеллекта / М.А. Кулагин, С.К. Хромов // Информатизация образования и науки. – 2019. – № 3. – С. 50-59.

3. Кулагин, М. А. Поддержка принятия решения по профилактике нарушений безопасности в работе / М.А. Кулагин, В.Г. Сидоренко // Надежность. – 2021. – Т. 21 – № 4. – С. 38-46.

4. Кулагин, М. А. Построение архитектуры интеллектуальной системы управления городской рельсовой транспортной системой / М.А. Кулагин, Л.А. Баранов, В.Г. Сидоренко, В.М. Алексеев // Мир транспорта. – 2021. – Т. 19 – № 1. – С. 18-46.

5. Кулагин, М. А. Подход к формированию рейтинга работы машиниста с использованием различных метрик сравнения / М.А. Кулагин, В.Г. Сидоренко // Электроника и электрооборудование транспорта. – 2018. – № 1. – С. 14-17.

б) в изданиях, индексируемых в Scopus:

6. Kulagin, M. Transport workers activities analysis using an artificial neural network / M. Kulagin, V. Sidorenko // International Conference on Intelligent Information Technologies for Industry. – 2018. – P. 308-316.

7. Kulagin, M. A Recommender Subsystem Construction for Calculating the Probability of a Violation by a Locomotive Driver using Machine-learning Algorithms / M. Kulagin, V. Sidorenko // 2020 IEEE East-West Design & Test Symposium (EWDTS). – 2020. – P. 1-5.

8. Kulagin, M. An Intelligent Evaluation System for Predicting Engine Driver Reliability / M. Kulagin, V. Sidorenko // Russian Electrical Engineering. – 2020. – Vol. 91 – № 9. – P. 587-591.

9. Kulagin, M. Machine learning application for support for automated control systems users / M. Kulagin, S. Khromov, V. Sidorenko // Journal of Physics: Conference Series. – 2020. – Vol. 1680 – № 1. – P. 12019.

10. Kulagin, M. Predicting the Failure of Traction Electric Motors of Electric Rolling Stock of Railways Using Deep Neural Networks / M. Kulagin, V. Sidorenko // Russian Electrical Engineering. – 2021. – Sep – Vol. 92 – № 9. – P. 515-519.

в) в официально зарегистрированных программах для ЭВМ:

11. Свидетельство о государственной регистрации программы для ЭВМ №2019614288 Россия. Доверенная среда локомотивного комплекса / Кулагин М.А., Харин О.В., Галицын М.А.; опубл. 26.04.2019.

12. Свидетельство о государственной регистрации программы для ЭВМ №2020612670 Россия. Автоматизированная система Доверенная среда локомотивного комплекса (2019) / Кулагин М.А., Харин О.В., Якимов С.М., Гоник М.М., Хлудеев М.А., Ярошук Д.И.; опубл. 23.03.2020.

г) в других изданиях и материалах конференций:

13. Кулагин, М. А. Детектирования железнодорожных светофоров с использованием нейронной сети / М.А. Кулагин // *DSPA: Вопросы применения цифровой обработки сигналов.* – 2017. – Т. 7 – № 4. – С. 67-71.

14. Кулагин, М. А. Прогнозирование совершения нарушения безопасности движения по вине локомотивной бригады с использованием современных методов машинного обучения / М.А. Кулагин, В.Г. Сидоренко // *Материалы XXVI международной научной конференции «Проблемы управления безопасностью сложных систем».* – 2018. – С. 318-323.

15. Кулагин, М. А. Разработка интеллектуальной системы сопровождения пользователей ИС ОАО «РЖД» с использованием нейронных сетей / М.А. Кулагин, В.Г. Сидоренко, С.К. Хромов // *Труды всероссийской научно-практической конференции «Неделя науки-2019».* –М.: РУТ(МИИТ). –2019.

16. Кулагин, М. А. Влияние человеческого фактора на безопасность движения поездов / М.А. Кулагин, В.Г. Сидоренко, А.В. Маркевич // *Материалы XXVII международной научной конференции «Проблемы управления безопасностью сложных систем».* – 2019. – С. 265-270.

17. Кулагин, М. А. Интеллектуальная система оценки и прогнозирования надёжности работы машиниста локомотива / М.А. Кулагин, В.Г. Сидоренко // *Электротехника.* – 2020. – № 9. – С. 71-75.

18. Кулагин, М. А. Прогнозирование выхода из строя тяговых электродвигателей электроподвижного состава железных дорог с использованием глубоких нейронных сетей / М.А. Кулагин, В.Г. Сидоренко // *Электротехника.* – 2021. – № 9. – С. 52-56.

19. Кулагин, М. А. Автоматизация сопровождения пользователей автоматизированных систем управления на базе машинного обучения / М.А. Кулагин, В.Г. Сидоренко, С.К. Хромов // *Новые информационные технологии в исследовании сложных структур.* – 2020. – С. 74-75.

20. Кулагин, М. А. Обеспечение безопасности применения речевых технологий в работе оперативного персонала городских рельсовых транспортных систем / М.А. Кулагин, В.Г. Сидоренко, Е.П. Балакина,

Л.Н. Логинова // Материалы XXIX международной научно-практической конференции «Проблемы управления безопасностью сложных систем». – 2021. – С. 355-361.

Кулагин Максим Алексеевич

**ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА АНАЛИЗА И
ПРОГНОЗИРОВАНИЯ НАРУШЕНИЙ ПРИ УПРАВЛЕНИИ
ПОДВИЖНЫМ СОСТАВОМ**

2.9.8. Интеллектуальные транспортные системы

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Подписано в печать «__» _____ 2022 Заказ № _____. Формат 60×84 1/16

Объем 1,5 усл. п.л.

Тираж 80 экз.