

# Построение маршрутов с учетом спрогнозированного риска дорожно-транспортных происшествий на основе исторических данных

А. Ф. Шаймарданов • Е. Ю. Сазонова • О. Н. Сметанина

Увеличение количества транспортных средств и водителей на дорогах ведет к росту количества дорожно-транспортных происшествий (ДТП). Одним из способов улучшения данной ситуации является внедрение интеллектуальных систем обработки транспортной информации, которые позволят эффективно управлять дорожным движением, выявлять потенциально опасные участки дорог, выбирать маршруты с учётом риска возникновения ДТП. В статье предлагается разработать решение поиска рациональных маршрутов с учетом оценки риска ДТП на основе исторических данных. Авторами приведены постановка задачи построения рациональных маршрутов, результаты анализа существующих программных решений в области построения маршрутов с учетом рисков составляющей на дорогах. Анализ позволил сделать вывод, что существующие решения предоставляют базовый функционал для анализа дорожной обстановки, но не предлагают детального анализа сегментов маршрута с учетом риска ДТП; большинство систем фокусируются на текущих данных о пробках и дорожных событиях, игнорируя долгосрочный анализ и прогнозирование. В работе решаются две задачи – прогнозирование риска ДТП и построение рационального маршрута с учетом выявленного риска. Для решения задачи прогнозирования риска была выбрана библиотека CatBoost, разработанная Яндексом. CatBoost отличается высокой эффективностью при работе с категориальными признаками и обладает встроенными средствами предотвращения переобучения. Для обучения модели машинного обучения был использован набор данных «Дорожно-транспортные происшествия: координаты, участники и пострадавшие». Для построения безопасного маршрута и визуализации его на карте используются OpenStreetMap в сочетании с Open Source Routing Machine. Предложенные решения имеют практическую ценность и могут служить основой для дальнейших исследований и развития интеллектуальных систем анализа транспортной информации.

*Построение маршрутов; прогнозирование риска ДТП; градиентный бустинг; CatBoost.*

## ВВЕДЕНИЕ

В современном мире безопасность дорожного движения является одной из приоритетных задач как для государственных органов, так и для частных транспортных компаний, страховых организаций и самих участников движения. С каждым годом увеличивается количество транспортных средств, что, в свою очередь, ведёт к росту количества дорожно-транспортных происшествий (ДТП). Эффективное управление дорожным движением, выявление потенциально опасных участков дорог, а также выбор маршрутов с учётом вероятности возникновения ДТП являются важнейшими направлениями в области интеллектуального анализа транспортной информации.

Развитие цифровых технологий и широкое распространение данных открывают новые возможности для анализа транспортных потоков. Современные методы машинного обучения позволяют выявлять скрытые закономерности в исторических данных о ДТП, предсказывать вероятность возникновения аварий и тем самым способствовать повышению безопасности

Шаймарданов А. Ф., Сазонова Е. Ю., Сметанина О. Н. Построение маршрутов с учетом спрогнозированного риска дорожно-транспортных происшествий на основе исторических данных // СИИТ. 2025. Т. 7, № 5(24). С. 109-123. DOI: [10.54708/2658-5014-SIIT-2025-no5-p109](https://doi.org/10.54708/2658-5014-SIIT-2025-no5-p109). EDN: TRRSBA.

Shaymardanov A. F., Sazonova E. Yu., Smetanina O. N. "Construction of routes taking into account the predicted risk of road accidents based on historical data" // SIIT. 2025. Vol. 7, no. 5(24), pp. 109-123. DOI: [10.54708/2658-5014-SIIT-2025-no5-p109](https://doi.org/10.54708/2658-5014-SIIT-2025-no5-p109). EDN: TRRSBA (In Russian).

на дорогах. Одним из перспективных направлений является создание интерактивных веб-приложений, которые предоставляют пользователю возможность не только построить маршрут между двумя точками, но и выбрать наиболее безопасный путь с учётом анализа статистических данных по аварийности.

Вопросы применения цифровых технологий, и в частности, анализ факторов на различные виды систем, представлены в работах [Сай25, Ско25], прогнозированием риска ДТП и анализом аварийности занимались авторы работ [Бан15, Куз20, Луц08, Овч12], вопросы оптимизации маршрутов и анализа рисков ДТП были рассмотрены в работах [Бур23, [Вал24], [Вал24], Гер23, Джу16], разработка интеллектуальных транспортных систем была обсуждена в [Box23, Инт24, Кап24, Ben25, Box23], этические нормы интеллектуальных транспортных системах были рассмотрены в [Beh25].

Несмотря на то, что данными вопросами занимаются отечественные и зарубежные специалисты, разработка приложения для построения рациональных маршрутов с учетом оценки риска ДТП на основе исторических данных о транспортных потоках повысит эффективность решения. Предлагаемая система может применяться для повышения безопасности в навигационных и логистических решениях, а также в качестве аналитического инструмента для выявления опасных участков дорожной сети. Реализация визуальной оценки рисков может способствовать снижению аварийности при соблюдении условий безопасного следования маршруту.

В статье отражены результаты анализа современного состояния проблемной ситуации, показаны постановка задачи построения маршрутов с учетом спрогнозированного риска ДТП; результаты аналитического обзора программных решений в области построения маршрутов; описан подход к определению риска ДТП на основе исторических данных и безопасного маршрута; представлены аспекты разработки программного решения.

### ПОСТАНОВКА ЗАДАЧИ ПОСТРОЕНИЯ МАРШРУТОВ С УЧЕТОМ СПРОГНОЗИРОВАННОГО РИСКА ДТП

Авторами предлагается разработать информационную систему, которая позволит построить несколько вариантов маршрутов между заданными точками; проанализировать каждый маршрут на предмет безопасности движения; разделить маршруты на сегменты и оценить уровень опасности для каждого сегмента; предоставить пользователю возможность выбора наиболее безопасного маршрута.

Разработка может быть разделена на 2 этапа: обработка исторических данных и построение безопасных маршрутов. Рассмотрим каждый этап подробнее.

*Этап 1.* Обработка исторических данных.

Входными данными являются координаты ДТП, сведения и обстоятельства этих ДТП (погода, дорожное покрытие, число пострадавших и т.д.). Выходными данными являются выявленные риски ДТП и вероятности.

*Этап 2.* Построение безопасных маршрутов.

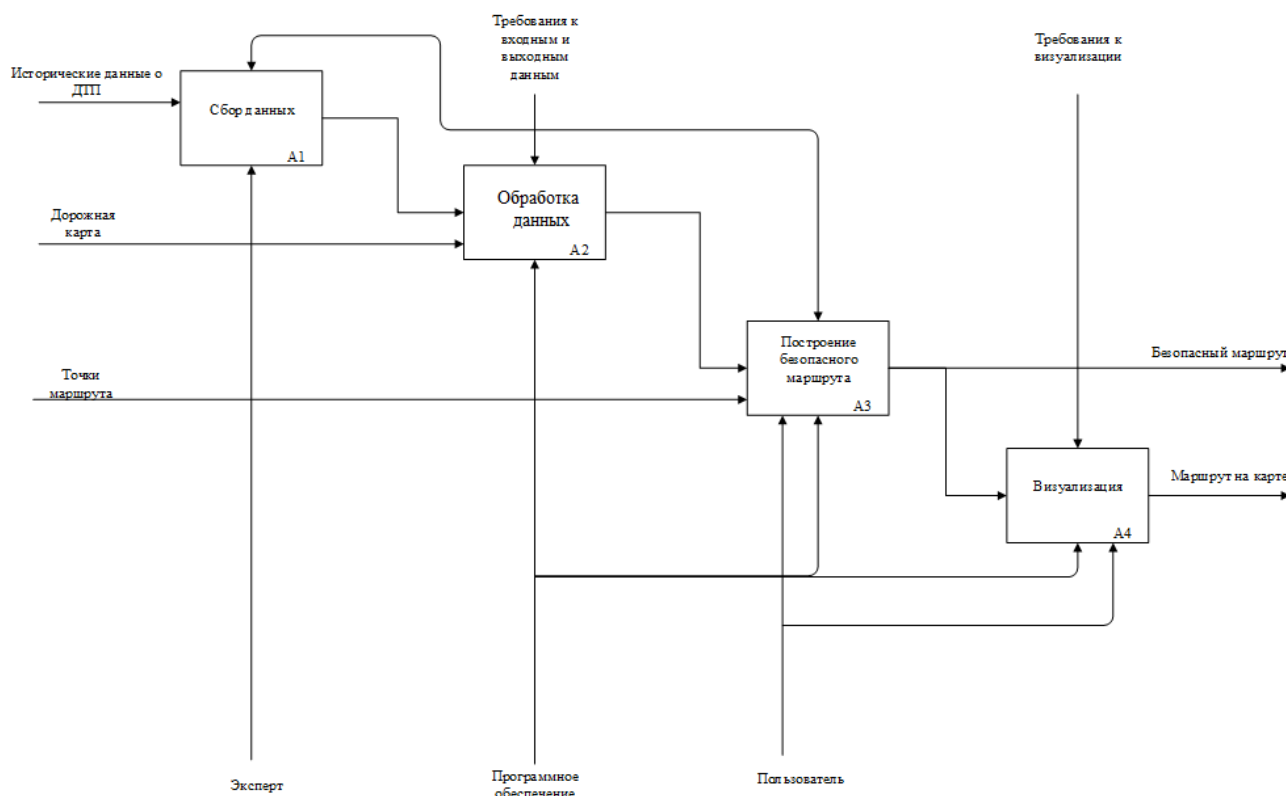
Входными данными выступают карта дорог; данные с рисками ДТП, основанные на происшествиях прошлых лет. Выходные данные – последовательность точек для посещения, визуализация маршрута на карте; отображение информации о рисках на выбранных маршрутах.

Таким образом, функциональные возможности программного обеспечения включают добавление и удаление точек маршрута пользователем; построение маршрута посегментно; визуализация риска ДТП на сегментах маршрута; визуализация риска ДТП на всём маршруте.

Структура решения задачи построения маршрутов с учетом спрогнозированного риска ДТП представлена на рис. 1.

В первом блоке необходимо определиться с характеристиками, описывающими ДТП и риск, собрать набор данных. На этапе обработки данных происходит очистка данных от пропущенных значений и ошибочных значений, преобразование бинарных и категориальных признаков, обучение и применение модели машинного обучения для вычисления баллов тяжести ДТП. На этапе построения безопасного маршрута высчитывается риск возникновения ДТП

и определение безопасного маршрута. На этапе визуализации пользователю предлагается решение в двух вариантах – маршрут, отображенный на карте (опасность сегментов выделяется разными цветами, наиболее безопасный маршрут выделяется среди прочих), текстовое описание (риск возникновения ДТП на маршрутах и сегментах описывается в процентах от 0 до 100).



**Рис. 1** Структура решения задачи построения маршрутов с учетом спрогнозированного риска ДТП

#### АНАЛИТИЧЕСКИЙ ОБЗОР ПРОГРАММНЫХ РЕШЕНИЙ В ОБЛАСТИ ПОСТРОЕНИЯ МАРШРУТОВ

Постоянный рост объемов транспортных перевозок и высокий уровень ДТП на дорогах требуют принятия мер по повышению безопасности дорожного движения. Одной из таких мер является применение информационных технологий в области анализа транспортного потока, построения маршрутов с учетом ситуации на дорогах. В настоящее время следующие решения могут использоваться для анализа транспортного потока: Яндекс.Карты, Google Maps, Безопасный город, Waze, 2ГИС. В табл. 1 представлен сравнительный анализ указанных решений. Рассмотрим более подробно отечественные решения.

Яндекс.Карты представляют собой ключевой инструмент пространственной аналитики и навигационного обеспечения. Платформа интегрирует многокомпонентную систему мониторинга транспортной инфраструктуры, основанную на обработке больших данных в режиме реального времени. Система способна к агрегации и анализу гетерогенных данных, обеспечивает оперативный учет изменений в дорожной сети, включая временные ограничения, связанные со строительными работами, а также динамическое отображение параметров трафика. Интеграция с сетью стационарных и мобильных датчиков позволяет детектировать такие события, как наличие автоматизированных систем фиксации нарушений ПДД, локальные скопления транспортных средств и аномалии движения. К преимуществам системы можно отнести обширную базу данных о дорожной ситуации, собранную на реальных данных пользователей;

обновлении информации в режиме реального времени; удобный интерфейс и интеграцию с другими сервисами Яндекса. К недостаткам системы относятся отсутствие детализированного анализа причин высокой аварийности на конкретных участках дороги; ограниченные возможности для прогнозирования вероятности ДТП на основе исторических данных; отсутствие персонализированных рекомендаций по выбору безопасного маршрута.

Таблица 1

### Сравнительный анализ программных решений

Критерий	Яндекс.Карты	Google Maps	Безопасный город	Waze	2ГИС
Географическое покрытие	РФ и страны СНГ	Глобальное	Ограничено крупными городами РФ	Глобальное	Основные города РФ
Актуальность данных	Высокая	Высокая	Высокая	Зависит от активности пользователей	Средняя
Прогноз ДТП	Отсутствует	Отсутствует	Частично	Частично	Отсутствует
Интеграция с данными	Пользовательские данные	Спутники и датчики	Государственные системы	Сообщество пользователей	Локальные данные
Анализ сегментов маршрута	Отсутствует	Отсутствует	Частично	Частично	Отсутствует
Удобство использования	Удобный	Удобный	Сложный	Удобный	Удобный

«Безопасный город» представляет собой интегрированный цифровой комплекс мониторинга и управления транспортной инфраструктурой, реализующий принципы предиктивной аналитики в сфере общественной безопасности. Система позволяет разрабатывать прогностические модели оценки аварийных рисков, основанных на мультиспектральном анализе пространственно-временных данных; автоматизировать детектирование аномалий дорожного движения, включая нарушения ПДД, нештатные ситуации и скопления транспорта. Алгоритмы идентифицируют паттерны, связывающие динамику транспортных потоков, историческую статистику ДТП, микроклиматические изменения и антропогенные факторы, формируя цифровые профили зон повышенной опасности. Предполагается, что внедрение подобных решений демонстрирует статистически значимое снижение показателей аварийности, а также сокращение экономических потерь, связанных с простоем транспорта и экологическим ущербом. К достоинствам данной системы можно отнести интеграцию с государственными системами видеонаблюдения; использование реальных данных о дорожной обстановке; широкое применение в регионах России. К недостаткам системы относится отсутствие интерактивного интерфейса для конечных пользователей; ограниченная доступность для частных лиц; низкая точность прогнозирования на основе статистических данных без использования сложных алгоритмов машинного обучения.

2ГИС функционирует как многоуровневая геоинформационная платформа, специализирующаяся на цифровой репрезентации городских пространств с акцентом на детализацию объектов инфраструктуры. Платформа реализует глубокую семантизацию картографического контента, а именно, каждый объект сопровождается многоаспектным профилем, включающим не только географические координаты, но и метаданные о потребительских характеристиках, исторических изменениях и социальной активности, что позволяет реализовывать предиктивную маршрутизацию, учитывающую как логистические параметры (протяженность, загруженность), так и контекстуальные факторы – доступность услуг, плотность пешеходных

потоков, сезонные паттерны спроса. Экосистема демонстрирует эволюцию картографических сервисов от навигационных инструментов к комплексным аналитическим платформам, где геоданные становятся основой для принятия стратегических решений в управлении городской средой. Достоинствами данной системы являются: подробная информация о городской инфраструктуре; эффективная работа в условиях крупных городов; простой и удобный интерфейс. Однако система имеет следующие недостатки: ограниченная функциональность для анализа дорожной безопасности; отсутствие глубокого анализа транспортного потока.

Проведенный анализ позволил сделать вывод, что существующие решения предоставляют базовый функционал для анализа дорожной обстановки, но не предлагают детального анализа сегментов маршрута с учетом риска ДТП; большинство систем фокусируются на текущих данных о пробках и дорожных событиях, игнорируя долгосрочный анализ и прогнозирование; системы, такие как «Безопасный город», имеют потенциал для анализа ДТП, но закрытый характер и ограниченное покрытие снижают их применимость для широкой аудитории. Таким образом, можно сделать вывод, что построение маршрутов с учетом спрогнозированной вероятности дорожно-транспортных происшествий на основе исторических данных является актуальной задачей, решение которой позволит предлагать пользователям безопасные маршруты, тем самым устранив недостатки существующих программных средств.

#### ОПРЕДЕЛЕНИЕ РИСКА ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ НА ОСНОВЕ ИСТОРИЧЕСКИХ ДАННЫХ

Рассмотрим математическую постановку задачи определения риска ДТП на дорогах. Дано: множество объектов (ДТП):  $x = \{x_1, x_2, \dots, x_n\}$ , где  $n$  – количество аварий; категориальные и бинарные признаки для каждого объекта; целевая переменная:  $severity\_score \in \{1, 3, 5\}$  (легкая, тяжелая и с погибшими).

Требуется найти функцию  $f(x)$ , которая предсказывает степень опасности ДТП:  $predicted\_risk = f(x)$  и преобразует предсказания в нормализованный риск  $risk\_normalized \in [0, 100]$ . Авторами учитываются следующие ограничения: категориальные признаки не могут быть напрямую использованы в числовых операциях; модель должна работать со смешанными типами данных (категории + числа); предсказанный риск должен быть интерпретируемым (шкала 0–100%).

Общий алгоритм построения модели для прогнозирования риска ДТП на основе исторических данных представлен далее:

*Шаг 1.* Предобработка и очистка данных: удаление пустых значений, выделение признаков, нормализация.

*Шаг 2.* Формирование выборки: разделение на обучающую и тестовую.

*Шаг 3.* Обучение модели: подбор внутренних параметров, минимизирующих ошибку.

*Шаг 4.* Оценка качества: проверка точности предсказаний, анализ ошибок.

*Шаг 5.* Применение: использование модели для прогнозов на новых данных.

В качестве основного источника о ДТП были взяты данные проекта «Карта ДТП» о ДТП в РФ с 2015 года [Дор24], преобразованных в формате плоских таблиц, в работе используется часть набора данных о ДТП с 2015 по 2024 год в Республике Башкортостан (рис. 2). В наборе рассматриваются следующие параметры ДТП: идентификатор ДТП; тип ДТП по наличию пострадавших пешеходов и детей; категория ДТП, например «Наезд на пешехода», «Опрокидывание», «Столкновение»; регион, в котором произошло ДТП; город или район; адрес; долгота; широта; объекты, которые находятся вблизи; дата и время ДТП; наличие дневного и искусственного освещения, например «В темное время суток, освещение отсутствует»; погодные условия, например «Ясно. Ураганный ветер», «Снегопад. Туман»; состояние дорожного покрытия; количество участников ДТП; категории участников ДТП, например «Пешеходы», «Велосипедисты», «Дети»; тяжесть ДТП по степени вреда здоровью, например «Легкий», «Тяжелый», «С погибшими»; количество погибших в ДТП; количество раненых в ДТП.



Общее количество наблюдений в наборе данных – 38 253. После предобработки набора данных, и именно, удаления пустых строк и ошибочных значений координат, количество наблюдений в наборе данных сократилось до 36 052.

№	Дата	Тип ДТП	Участники	Место	Адрес	Координаты	Степень тяжести	Исход
1	2016-06-01	Дорожно-транспортные происшествия	Наезд на велосипедиста	Республика Башкортостан	Ижевский	г. Ижевск, ул. Некрасова, 21а	54.942923	56.272134 Многоквартирные жилые дома, Нерегулируемый пешеходный переход
2	2016-06-01	Дорожно-транспортные происшествия	Съезд с дороги	Республика Башкортостан	Ижевский	г. Ижевск, ул. Яковлева, 43	54.914946	56.288991 Производственное предприятие
3	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	г. Ижевск, ул. Советская, 4	54.934935	56.268411 Нерегулируемый пешеходный переход
4	2016-06-01	Дорожно-транспортные происшествия	Опронидывание	Республика Башкортостан	Ижевский	с. Кузино, ул. Садовая, 13	54.931165	56.203583 Жилые дома индивидуальной застройки
5	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, АБС/Белая - Ама, 9 км	54.94345	56.226546 Неизвестно
6	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, ул. Некрасова, 21	54.943056	56.271991 Многоквартирные жилые дома, Нерегулируемый пешеходный переход
7	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, ул. Станционная, 15	54.930192	56.271354 Автоостановка (отделенная от проезжей части), Аэропорт, м/д вокзал (м/д станций), с
8	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 46 км	54.957347	56.281080 Жилые дома индивидуальной застройки, Нерегулируемый перекресток неравнознач
9	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	г. Ижевск, пер. Маяковского, 11	54.927424	56.259056 Жилые дома индивидуальной застройки, Нерегулируемый перекресток неравнознач
10	2016-06-01	Дорожно-транспортные происшествия	Наезд на велосипедиста	Республика Башкортостан	Ижевский	г. Ижевск, Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 47 км	54.929012	56.280098 Жилые дома индивидуальной застройки, Нерегулируемый перекресток неравнознач
11	2016-06-01	Дорожно-транспортные происшествия	Опронидывание	Республика Башкортостан	Ижевский	Ижевск - Ижевск - Нават, 7 км	54.954214	56.204412 Неизвестно
12	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 20 км	54.627092	56.182505 Неизвестно
13	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 45 км	54.960877	56.278573 Нерегулируемый перекресток неравнозначных улиц (дорог), Нерегулируемый пере
14	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Уфа - Бирск - Ижевск (Благовещенск, Бураево, Краснохолмский), 185 км	55.093216	56.041867 Неизвестно
15	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 49 км	54.954257	56.268447 Нерегулируемый перекресток неравнозначных улиц (дорог), АЗС
16	2016-06-01	Дорожно-транспортные происшествия	Опронидывание	Республика Башкортостан	Ижевский	Мескунто - д/д Четырман - Берный Чат, 0 км	55.358198	56.09172 Жилые дома индивидуальной застройки
17	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	г. Ижевск, Уфа - Бирск - Ижевск (Благовещенск, Бураево, Краснохолмский), 210 км	54.972121	56.249997 АЗС
18	2016-06-01	Дорожно-транспортные происшествия	Съезд с дороги	Республика Башкортостан	Ижевский	Ижевск - Старый Кудым, 7 км	54.917513	56.271598 Неизвестно
19	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 41 км	54.934924	56.249836 Жилые дома индивидуальной застройки, Нерегулируемый перекресток неравнознач
20	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Ижевск - Старый Кудым, 26 км	55.242691	56.245198 Неизвестно
21	2016-06-01	Дорожно-транспортные происшествия	Съезд с дороги	Республика Башкортостан	Ижевский	Ижевск - ул. Победы, 60	54.92448	56.26888 Многоквартирные жилые дома
22	2016-06-01	Дорожно-транспортные происшествия	ДТП и п. Наезд на пешехода	Республика Башкортостан	Ижевский	Уфа - Бирск - Ижевск (Благовещенск, Бураево, Краснохолмский), 204 км	55.03001	56.20186 Производственное предприятие
23	2016-06-01	Дорожно-транспортные происшествия	ДТП и п. Стопкование	Республика Башкортостан	Ижевский	м/д Ижевск - Верхние Татышлы - Шуганово - Янгиль - Нават - Майск, 5 км	55.162868	56.150802 Неизвестно
24	2016-06-01	Дорожно-транспортные происшествия	ДТП и п. Стопкование	Республика Башкортостан	Ижевский	с. Каймашбаши, ул. Школьная, 7	54.917778	56.261667 Жилые дома индивидуальной застройки, Нерегулируемый перекресток равнознач
25	2016-06-01	Дорожно-транспортные происшествия	ДТП и п. Наезд на велосипедиста	Республика Башкортостан	Ижевский	г. Ижевск, ул. Амина, 29	54.9281	56.2644 Выезд с прилегающей территории, Административный здание
26	2016-06-01	Дорожно-транспортные происшествия	ДТП и п. Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, ул. Ромашовская, 21	54.9284	56.2631 Многоквартирные жилые дома, Нерегулируемый пешеходный переход, Школа либ
27	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 26 км	54.7044	56.2019 Неизвестно
28	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, ул. Советская, 12	54.9557	56.2697 Многоквартирные жилые дома, Нерегулируемый пешеходный переход
29	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Уфа - Бирск - Ижевск (Благовещенск, Бураево, Краснохолмский), 211 км	54.974683	56.244893 Неизвестно
30	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 20 км	54.627124	56.182505 Жилые дома индивидуальной застройки
31	2016-06-01	Дорожно-транспортные происшествия	Опронидывание	Республика Башкортостан	Ижевский	д. Тартар, ул. Центральная, 32	54.6522	56.2197 Жилые дома индивидуальной застройки
32	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Ижевск - Старый Кудым, 0 км	55.0142	56.2730 Жилые дома индивидуальной застройки, Нерегулируемый перекресток неравнознач
33	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, ул. Советская, 20	54.9378	56.2717 Многоквартирные жилые дома
34	2016-06-01	Дорожно-транспортные происшествия	Стопкование	Республика Башкортостан	Ижевский	Нефтекамск - Ижевск (Кувакино, Сутырьево, Абулган), 20 км	54.93141	56.26855 Неизвестно
35	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, ул. Победы, 78	54.934443	56.270686 Многоквартирные жилые дома, Нерегулируемый пешеходный переход
36	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	г. Ижевск, ул. Победы, 50	54.9235	56.2678 Многоквартирные жилые дома
37	2016-06-01	Дорожно-транспортные происшествия	Наезд на пешехода	Республика Башкортостан	Ижевский	Ижевск - д. Татышлы (Сандуки, Верхнекудашево), 11 км	55.0736	56.155 Неизвестно

Рис. 2 Фрагмент набора данных о ДТП

Для решения задачи предсказания вероятности дорожно-транспортного происшествия на основе исторических данных могут быть применены следующие методы: логистическая регрессия, случайный лес, градиентный бустинг (XGBoost), CatBoost (табл. 2).

Таблица 2

Сравнительный анализ методов определения риска ДТП

Критерий	Логистическая регрессия	Случайный лес	XGBoost	CatBoost
Поддержка категориальных признаков	Требуется кодирование	Ограниченная	Требуется кодирование	Встроенная
Устойчивость к выбросам	Низкая	Высокая	Высокая	Высокая
Работа с большим числом признаков	Средняя	Хорошая	Отличная	Отличная
Точность на больших данных	Низкая	Хорошая	Хорошая	Отличная
Параметры для настройки	Минимальные	Средние	Много	Мало

Логистическая регрессия является базовым алгоритмом классификации, используемым для моделирования вероятности бинарного события, в текущей работе таким событием будет «наступление ДТП». Алгоритм легко интерпретируется и быстро обучается, но ограничен линейной зависимостью между признаками и целевой переменной, что снижает его эффективность в задачах с комплексной структурой данных. Таким образом, логистическая регрессия может быть использована для прогнозирования вероятности ДТП на определенном участке дороги, учитывая такие признаки, как интенсивность движения, погодные условия, время суток и наличие пешеходов.

В основу следующих алгоритмов положены основные принципы построения деревьев решений. Случайный лес относится к ансамблевым алгоритмам, представляющим собой множество деревьев решений, построенных на различных подвыборках обучающего набора. Алгоритм работает следующим образом: создается набор случайных подвыборок обучающего набора данных (обычно с возвращением, метод bagging); для каждой подвыборки строится дерево решений; предсказание делается путем голосования (для классификации). Алгоритм

может эффективно работать с шумами, однако сложен в интерпретации и может быть ресурсоёмким при большом числе деревьев. Алгоритм не требует масштабирования данных, работает с нелинейными зависимостями, имеет высокую точность благодаря ансамблевому подходу. Модель, построенная с помощью данного алгоритма, может прогнозировать тяжесть аварии на основе исторических данных, учитывая множество факторов, таких как погода, состояние дороги и тип транспортного средства. Градиентный бустинг – это метод построения ансамбля моделей, где каждая последующая модель исправляет ошибки предыдущих. XGBoost (eXtreme Gradient Boosting) является оптимизированным бустингом, обладающим высокой точностью, но чувствительным к гиперпараметрам. Алгоритм начинается с простой модели, и на каждом шаге обучается новая модель, которая минимизирует ошибки предыдущих моделей, итоговым предсказанием выступает взвешенная сумма предсказания всех моделей. Алгоритм хорошо работает на структурированных данных, но требует ручного кодирования категориальных признаков. XGBoost может использоваться для прогнозирования риска ДТП с учётом множества факторов, таких как географическое расположение, время года и дорожные условия. CatBoost является также реализацией градиентного бустинга, ключевой особенностью является встроенная поддержка категориальных признаков.

Таким образом, можно сделать вывод, что наилучшим для определения риска ДТП на основе исторических данных является CatBoost, разработанный компанией Яндекс [Cat17]. Одним из достоинств метода является возможность работы с категориальными признаками напрямую, без необходимости кодирования (One-Hot, Label Encoding). Это является особенно важным для рассматриваемой задачи, так как количество признаков (например, «погода», «категории участников», «тип освещения») являются категориальными. Метод имеет повышенную устойчивость к переобучению, то есть благодаря использованию принципов случайности при построении деревьев CatBoost стабильно показывает хорошее качество даже при ограниченном объёме данных. Модель, построенная с помощью CatBoost, имеет высокую точность при приемлемом времени обучения, что позволяет легко применять и заново обучать ее при необходимости. При реализации веб-приложений важно, чтобы модель эффективно внедрялась, модель в продакшн-коде CatBoost может быть сохранена в формате \*cbm и легко быть использована в Python-проектах без сложной инфраструктуры.

Рассмотрим пошаговый алгоритм CatBoost:

*Шаг 1. Подготовка данных.* Категориальные признаки (погода, освещение) автоматически обрабатываются без ручного кодирования. Бинарные признаки (наличие пешеходов, детей) преобразуются в числовые (0 или 1). Целевая переменная (тяжесть ДТП) кодируется числами: 1 (лёгкое), 3 (тяжёлое), 5 (с погибшими).

*Шаг 2. Построение деревьев решений.* CatBoost строит ансамбль деревьев (например, 300 деревьев). Каждое дерево обучается на остатках ошибок предыдущих деревьев (градиентный бустинг). Учитывает категориальные признаки без предварительного one-hot кодирования.

*Шаг 3. Борьба с переобучением.* Используется ранняя остановка: обучение прекращается, если ошибка на тестовых данных перестаёт уменьшаться. Ограничивается глубина деревьев (например,  $\text{depth} = 4$ ), чтобы избежать сложных зависимостей. Применяется стохастическое градиентное бустинг (шаг обучения  $\text{learning\_rate} = 0.2$ ).

*Шаг 4. Предсказание и оценка.* Модель предсказывает уровень риска для каждого ДТП. Оценивается качество через RMSE (среднеквадратичная ошибка). Риск нормализуется к шкале 0–100% для удобства интерпретации.

*Шаг 5. Экспорт результатов.* Данные сохраняются в формате JSON (координаты + риск) для визуализации на карте.

Функция предсказания выглядит следующим образом:  $\text{predicted\_risk}_i = \sum_{m=1}^M h_m(x_i)$ , где  $h_m(x_i)$  – предсказание  $m$ -го дерева для объекта  $x_i$  и  $M$  – общее количество деревьев. На первой итерации ( $m = 1$ ):  $h_1(x_i)$  обучается на исходных данных ( $\text{severity\_score}$ ).

На последующих итерациях ( $m > 1$ ): каждое новое дерево  $h_m$  обучается на остатках (разнице между истинным значением и предсказанием предыдущих деревьев):

$$r_i^{(m)} = \text{severity\_score}_i - \sum_{k=1}^{m-1} h_k(x_i);$$

темп обучения  $\eta$  замедляет переобучение:

$$F_m(x_i) = F_{m-1}(x_i) + \eta h_m(x_i);$$

нормализованный риск:

$$\text{risk\_normalized}_i = \frac{\text{predicted\_risk}_i - R_{\min}}{R_{\max} - R_{\min}} \cdot 100,$$

где

$$R_{\min} = \min(\text{predicted\_risk}_1, \dots, \text{predicted\_risk}_n);$$

$$R_{\max} = \max(\text{predicted\_risk}_1, \dots, \text{predicted\_risk}_n).$$

Авторами был проведен вычислительный эксперимент с целью выявления наилучших параметров алгоритма. Сравнительные результаты пяти различных конфигураций представлены в табл. 3, другие конфигурации показали близкие значения, что подтверждает стабильность модели. В качестве метрик для оценки конфигураций были использованы две метрики: RMSE (Root Mean Squared Error) и MAE (Mean Absolute Error). Наилучшей конфигурацией является  $\text{iterations} = 300$  (количество деревьев, которые влияют на переобучение),  $\text{depth} = 4$  (глубина деревьев, которая отвечает за сложность модели),  $\text{learning\_rate} = 0.2$  (шаг обучения, который влияет на скорость сходимости),  $\text{RMSE} = 1.28$  означает, что в среднем ошибка предсказания – около 1.28 балла тяжести ДТП,  $\text{MAE} = 1.1$  означает, что в среднем ошибка  $\approx 1$  балл тяжести.

Таблица 3

Сравнительный анализ конфигураций модели

Параметры алгоритма	RMSE	MAE
$\text{iterations} = 500, \text{depth} = 5, \text{learning\_rate} = 0.1$	1.2837	1.1161
$\text{iterations} = 1000, \text{depth} = 6, \text{learning\_rate} = 0.05$	1.2844	1.1164
$\text{iterations} = 1000, \text{depth} = 6, \text{learning\_rate} = 0.05$	1.2847	1.1165
$\text{iterations} = 300, \text{depth} = 4, \text{learning\_rate} = 0.2$	1.2826	1.1161
$\text{iterations} = 1000, \text{depth} = 5, \text{learning\_rate} = 0.07$	1.2843	1.1171

### ОПРЕДЕЛЕНИЕ МАРШРУТА

#### С УЧЕТОМ РИСКА ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ

Безопасный маршрут между двумя точками определяется с учетом найденного риска ДТП на предыдущем этапе решения задачи. Важной составляющей реализацией программного решения является сервис для визуализации маршрута. Авторами проанализированы два популярных сервиса для анализа и визуализации – OpenStreetMap (OSM) и Яндекс.Карты API. OpenStreetMap – это открытый проект, предоставляющий свободные геоданные под лицензией Open Database License (ODbL), проект основан на принципах краудсорсинга, где пользователи могут добавлять, редактировать и обновлять данные о картах. К достоинствам данного сервиса можно отнести, что данные доступны бесплатно и могут быть использованы без ограничений; имеется возможность загрузки полных дампов данных для локальной обработки; сервис подходит для проектов любого уровня сложности благодаря отсутствию ограничений на объем данных; имеется активное сообщество разработчиков и энтузиастов, которые поддерживают актуальность данных. Однако качество данных может варьироваться в зависимости от региона,



так как проект основан на вкладе пользователей, поэтому требуется дополнительная работа по преобразованию данных в удобный формат для использования. Таким образом, необходимо самостоятельно реализовывать функционал, такой как маршрутизация и анализ пробок.

Яндекс.Карты API – это коммерческий сервис, предоставляемый компанией Яндекс, предлагает широкий набор инструментов для работы с картами, включая маршрутизацию, анализ пробок и другие функции. Достоинствами сервиса являются следующие аспекты: API предоставляет готовые функции для маршрутизации, анализа пробок и других задач; данные регулярно обновляются и поддерживаются профессиональной командой; возможность использования дополнительных сервисов, таких как Яндекс.Навигатор и Яндекс.Такси. Однако нужно учитывать недостатки данного сервиса при разработке программного решения: бесплатная версия имеет ограничения на количество запросов, что может быть критично для крупных проектов; доступ к данным ограничен, и их использование возможно только через API; проект становится зависимым от политики компании Яндекс.

Сравнительный анализ сервисов представлен в табл. 4.

Таблица 4

### Сравнительный анализ сервисов OpenStreetMap и Яндекс.Карты API

Критерии	OSM	Яндекс.Карты API
Доступность	Открытая лицензия и бесплатный доступ	Закрывающаяся лицензия и ограниченный доступ по запросу
Качество данных	Варьируется, зависит от региона	Высокое, постоянные обновления
Поддержка	Нет официальной поддержки	Официальная поддержка
Стоимость	Бесплатно	Бесплатно, но с ограничениями
Сложность построения маршрутов	Сложная настройка, большее отклонение идет в сторону Linux	Простая настройка, универсально

На основе проведенного анализа можно сделать вывод, что среди рассмотренных решений OpenStreetMap в сочетании с OSRM (библиотека планирования маршрутов) обеспечивает полную свободу действий, открытые данные, удобную интеграцию с веб-интерфейсом. Все перечисленное делает его идеальным выбором для академической разработки веб-приложения, где ключевыми являются доступность, гибкость и расширяемость. Несмотря на отсутствие поддержки онлайн-пробок, система может эффективно использовать исторические данные ДТП для моделирования риска, что делает её концептуально сопоставимой с платными коммерческими платформами.

Построение маршрутов и визуализация происходят следующим образом: для двух точек, выбранных пользователем, маршруты строятся с использованием OSRM, каждый маршрут разбивается на сегменты; в свою очередь для каждого сегмента по его географическому положению выбираются ближайшие точки из набора с риском ДТП; риск сегмента рассчитывается как среднее значение риска ближайших точек (в радиусе 200 метров); итоговый риск всего маршрута считается, как усреднение по всем сегментам; маршруты визуализируются цветными линиями, где цвет отражает уровень риска (зелёный – безопасная часть маршрута, красный – опасная часть маршрута). Радиус в 200 метров был выбран экспериментально как компромисс, такой выбор позволяет захватить локальные участки дороги, а не район целиком; при этом не размывает риски за счёт дальних точек; адекватно соотносится с городской застройкой (примерно один квартал). Таким образом, каждый сегмент маршрута проверяется на наличие «опасных точек» вблизи, и ему присваивается риск, соответствующий среднему риску этих ближайших происшествий. Для определения того, какие участки маршрута проходят через опасные зоны, в работе используется анализ ближайших точек риска, в частности, для каждой

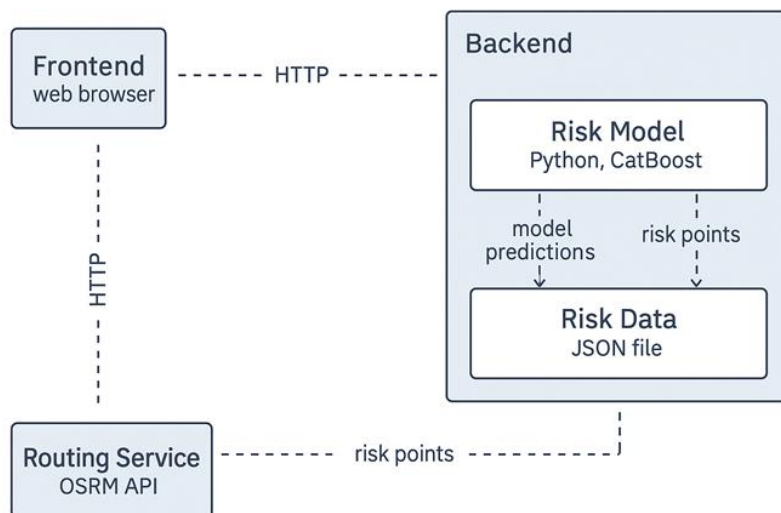
точки маршрута вычисляется расстояние до ближайших точек с историческими ДТП; если в пределах заданного радиуса находятся точки с высоким риском – это влияет на цвет сегмента маршрута; итоговая оценка маршрута формируется как среднее значение рисков вдоль всей траектории.

Такой подход позволяет пользователю не просто построить кратчайший путь, а выбрать рациональный маршрут по критерию безопасности, учитывая различные факторы.

### ПРОГРАММНОЕ РЕШЕНИЕ ДЛЯ ОПРЕДЕЛЕНИЯ МАРШРУТА С УЧЕТОМ РИСКА ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ

В процессе проектирования программного обеспечения особое внимание было уделено выбору программно-технического стека, обеспечивающего надёжную реализацию поставленной задачи – построение маршрутов с учетом риска дорожно-транспортных происшествий на основании исторических данных. С учётом характера задачи и требований к производительности, визуализации и точности прогнозирования был сформирован следующий стек технологий: язык программирования проекта Python 3.12; библиотеки для обработки табличных и числовых данных Pandas и NumPy, которые применялись на этапе чтения, очистки и подготовки признаков ДТП, в том числе для генерации дополнительных признаков на основе текстовых столбцов; CatBoost – фреймворк градиентного бустинга, который был использован для предсказания риска ДТП на основании характеристик происшествия; Scikit-learn использовался для разделения выборки на обучающую и тестовую, расчёта метрики RMSE (среднеквадратичная ошибка), необходимой для оценки точности модели; библиотека Leaflet.js – для отображения веб-карт, OpenStreetMap (OSM) применена в качестве основного поставщика картографических данных, OSRM (Open Source Routing Machine) – открытое решение для построения маршрутов на основе географических координат; Nominatim API – открытый API для прямого и обратного геокодирования, задействован был для обеспечения преобразования текстовых адресов в координаты и обратно, используется для автодополнения и отображения адреса при установке точек маршрута вручную.

Диаграмма развертывания программного решения (рис. 3) показывает взаимосвязь узлов системы, каналов связи между ними и размещения программных файлов на аппаратном обеспечении.

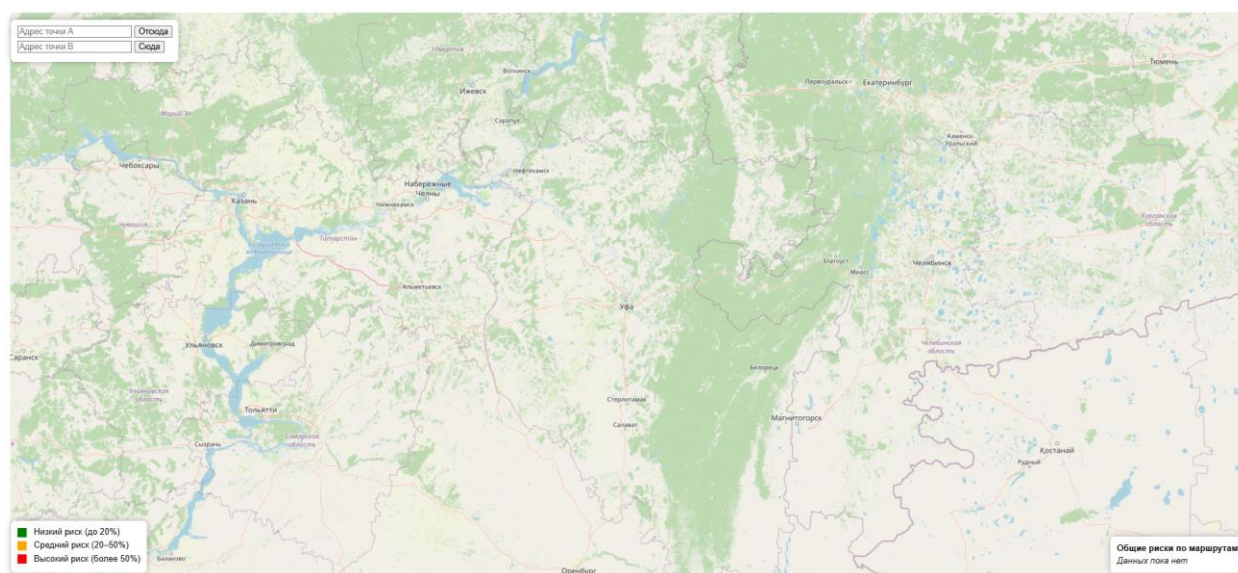


**Рис. 3** Диаграмма развертывания программного решения для определения маршрута с учетом риска ДТП

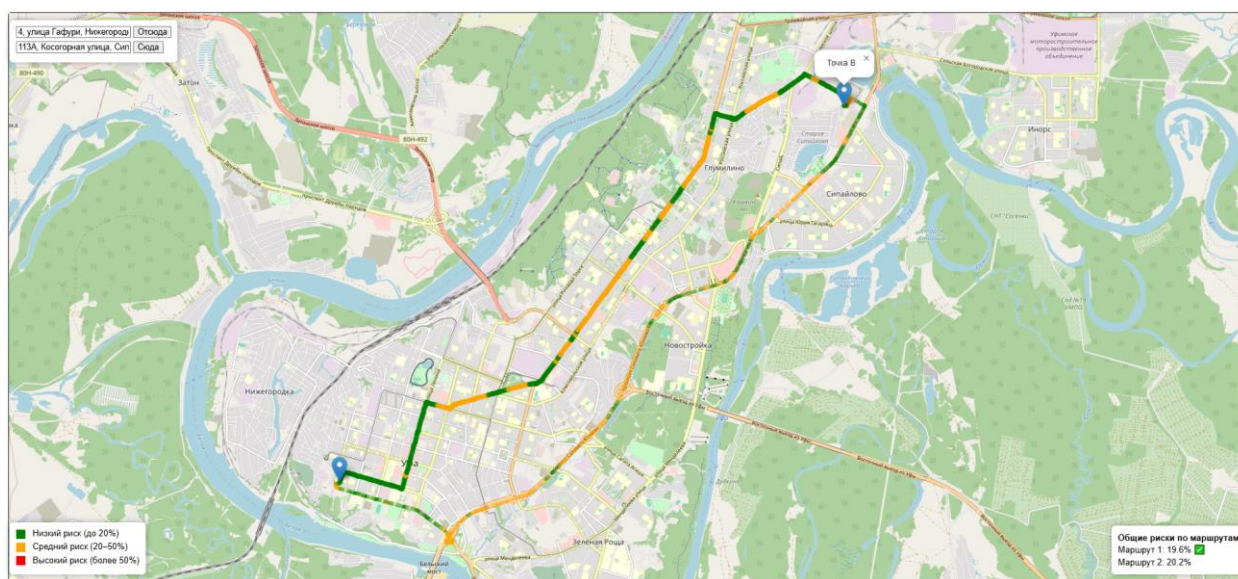
Архитектура программного решения реализована в виде двухэтапного подхода, что обеспечивает удобство и независимость клиентской части от серверной логики. На этапе анализа

данных осуществляются разовая предобработка и обучение модели машинного обучения, данный процесс реализуется на языке Python в виде отдельного скрипта `train_model.py`. Задачами являются загрузка и очистка данных; построение признаков; обучение модели `CatBoostRegressor`; расчёт вероятности риска ДТП для каждой записи; нормализация значений в диапазоне от 0 до 100; сохранение результата в файл `dtg_points_risk.json`. После завершения этапа данные готовы для использования на клиенте без необходимости повторного запуска анализа при каждом действии пользователя. Второй этап представляет собой интерактивную веб-страницу, работающую локально в браузере. Веб-интерфейс создан на базе библиотеки `Leaflet.js` и использует: HTML, CSS и JavaScript; библиотеку `Leaflet` для визуализации карты и маршрутов; `Nominatim API` – для геокодирования и автоподстановки адресов; `OSRM API` – для построения до трёх маршрутов между двумя точками; локальный JSON-файл с рисками (`dtg_points_risk.json`) – для анализа и визуализации риска по маршрутам. Приложение не требует подключения к серверу, базам данных или сторонним back-end API. Весь функционал работает в браузере пользователя и использует только открытые публичные API и заранее подготовленные локальные данные.

Интерфейс пользователя представлен на рис. 4.

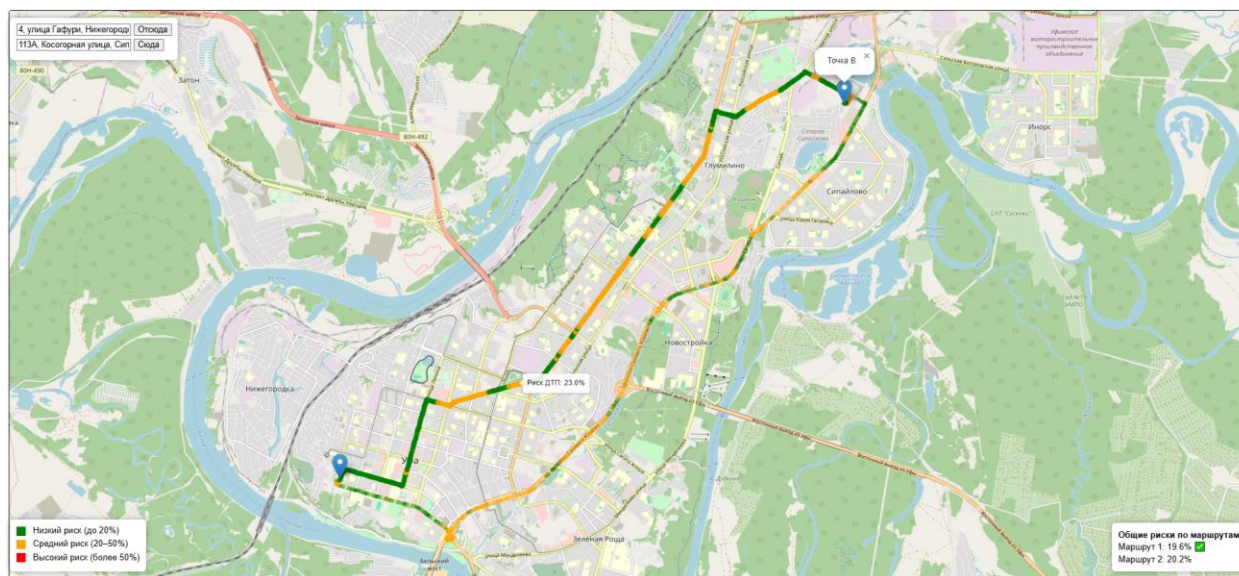


а



б





6

**Рис. 4** Интерфейс пользователя:

*а – основное окно приложения; б – отображение выбранных маршрутов;  
в – отображение вероятности на сегменте маршрута*

На основном экране пользователя отображаются легенда обозначения цветов у сегментов; блок, где отображаются текущие маршруты с вероятностями ДТП на пути следования; блоки для ввода адресов (рис. 4, а). Пользователь может выбрать кликом по карте две точки на карте или ввести через адресные строки, обозначающие начало и конец маршрута соответственно, отображаются маршруты, в блоке адресов отображаются вероятности ДТП на пути следования маршрута, на самой карте наиболее безопасный маршрут отрисован толстой и сплошной линией, остальные маршруты показаны более блекло и пунктирной линией (рис. 4, б). При наведении на любые сегменты маршрутов пользователь может увидеть риск ДТП на конкретном участке маршрута (рис. 4, в).

## ЗАКЛЮЧЕНИЕ

В работе рассмотрены ключевые аспекты, касающиеся анализа транспортных потоков и оценки дорожной безопасности с применением современных информационных технологий. Авторами сформулирована задача по разработке системы, которая будет учитывать статистически подтвержденные данные о риске ДТП на каждом участке. Было предложено разделить разработку программного решения на 2 этапа: обработка исторических данных и построение безопасных маршрутов. На этапе анализа данных осуществляются разовая предобработка и обучение модели машинного обучения. В качестве методов для обучения модели машинного обучения были рассмотрены логистическая регрессия, случайный лес, градиентный бустинг (XGBoost), CatBoost. Наилучшей альтернативой оказался CatBoost, так как имеются встроенная поддержка категориальных признаков, устойчивость к переобучению, высокая точность и скорость без сложности настройки гиперпараметров. Для построения безопасного маршрута и визуализации его на карте используются OpenStreetMap в сочетании с Open Source Routing Machine.

Практическая значимость исследования состоит в том, что использование разработанного программного обеспечения позволяет не только строить рациональные маршруты, но и учитывать риск ДТП на каждом участке пути. Система может применяться для повышения безопасности в навигационных и логистических решениях, а также в качестве аналитического инструмента для выявления опасных участков дорожной сети. Реализация визуальной оценки рисков может способствовать снижению аварийности при соблюдении условий безопасного

следования маршруту. Научная новизна исследования представлена разработанным алгоритмом оценки риска ДТП с использованием модели машинного обучения CatBoost на основе исторических данных.

## СПИСОК ЛИТЕРАТУРЫ | REFERENCES

- [Cat17] CatBoost: библиотека градиентного бустинга от Яндекса. 2017. URL: <https://catboost.ai>.
- [Beh25] Behki P., Shreya V. et al. Ethical Concerns of Autonomous Vehicles: An AI Framework // СИИТ. 2025. Т. 7, № 3(22). С. 20–28. DOI: [10.54708/2658-5014-SIIT-2025-no3-p20](https://doi.org/10.54708/2658-5014-SIIT-2025-no3-p20). EDN: HHYJIO.
- [Ben25] Benny B., Chellaswamy K. P. AI's Impact on Vehicle Growth and Ownership Patterns in the Automotive Industry // СИИТ. 2025. Т. 7, № 3(22). С. 40–47. DOI: [10.54708/2658-5014-SIIT-2025-no3-p40](https://doi.org/10.54708/2658-5014-SIIT-2025-no3-p40). EDN: QCQXLS.
- [Бан15] Банушкина Н. А., Печатнова Е. В. Повышение эффективности прогнозирования ДТП на автомобильных дорогах вне населенных пунктов на основе разработки экспертной системы // Известия Алтайского государственного университета. 2015. № 1/2(85). С. 86–90. DOI: [10.14258/izvasu\(2015\)1.2-15](https://doi.org/10.14258/izvasu(2015)1.2-15). EDN: UMGUQF.
- [Бур23] Буряковский П. К. Подход к определению кратчайшего расстояния между объектами разнородной геопространственной геометрии // СИИТ. 2023. Т. 5, № 5(14). С. 88–94. DOI: [10.54708/2658-5014-SIIT-2023-no5-p88](https://doi.org/10.54708/2658-5014-SIIT-2023-no5-p88). EDN: ZWGBWG.
- [Вал24] Валеев Р. С., Юсупов А. Р., Кашаев Ш. З., Валеева А. Ф. Об одной задаче обеспечения эффективного товарооборота дистрибьютерной компании // СИИТ. 2024. Т. 6, № 4(19). С. 38–50. DOI: [10.54708/2658-5014-SIIT-2024-no4-p38](https://doi.org/10.54708/2658-5014-SIIT-2024-no4-p38). EDN: KSRBGK.
- [Вох23] Вохминцев А. В. Методология решения проблемы одновременной навигации и построения карты на основе комбинирования визуальных и семантических характеристик окружающей среды // СИИТ. 2023. Т. 5, № 3(12). С. 136–155. DOI: [10.54708/2658-5014-SIIT-2023-no4-p136](https://doi.org/10.54708/2658-5014-SIIT-2023-no4-p136). EDN: EFOFFO.
- [Гер23] Герштейн А. М., Терехов А. Н. Обход опасных участков маршрута как способ повышения безопасности движения (на примере Санкт-Петербурга) // Компьютерные инструменты в образовании. 2023. № 1. С. 30–39. DOI: [10.32603/2071-2340-2023-1-30-39](https://doi.org/10.32603/2071-2340-2023-1-30-39). EDN: LBMCDH.
- [Джу16] Джурук Д. С. Методика повышения безопасности дорожного движения на двухполосных дорогах в местах концентрации ДТП на примере Сибирского федерального округа: дисс. ... канд. техн. наук. 2020. 176 с. EDN: KCQGVV.
- [Дор24] Дорожно-транспортные происшествия: координаты, участники и пострадавшие // Если быть точным. 2024. URL: [https://tochno.st/datasets/traffic\\_accidents](https://tochno.st/datasets/traffic_accidents).
- [Инт24] Интеллектуальные транспортные системы: из 2024 в 2025: Интернет-ресурс аналитического агентства TAdviser. [https://www.tadviser.ru/index.php/Статья:Интеллектуальные\\_транспортные\\_системы:\\_из\\_2024\\_в\\_2025?erid=2W5zFJaCHdr](https://www.tadviser.ru/index.php/Статья:Интеллектуальные_транспортные_системы:_из_2024_в_2025?erid=2W5zFJaCHdr)
- CatBoost: Gradient boosting library from Yandex. URL: <https://catboost.ai>.
- Behki P., Shreya V. et al. "Ethical Concerns of Autonomous Vehicles: An AI Framework" // SIIT. 2025. V. 7, № 3(22), pp. 20–28. DOI: [10.54708/2658-5014-SIIT-2025-no3-p20](https://doi.org/10.54708/2658-5014-SIIT-2025-no3-p20). EDN: HHYJIO.
- Benny B., Chellaswamy K. P. "AI's Impact on Vehicle Growth and Ownership Patterns in the Automotive Industry" // SIIT. 2025. V. 7, № 3(22), pp. 40–47. DOI: [10.54708/2658-5014-SIIT-2025-no3-p40](https://doi.org/10.54708/2658-5014-SIIT-2025-no3-p40). EDN: QCQXLS.
- Banushkina N. A., Pechatnova E. V. "Improving the efficiency of traffic accidents prediction on roads outside of settlements on the basis of an expert system development" // Bulletin of the Altai State University. 2015. . № 1/2(85). P. 86–90. (In Russian). DOI: [10.14258/izvasu\(2015\)1.2-15](https://doi.org/10.14258/izvasu(2015)1.2-15). EDN: UMGUQF.
- Buryakovsky P. K. "An approach to determining the shortest distance between objects of heterogeneous spatial geometry" // SIIT. 2023. V. 5, № 5(14), pp. 88–94. DOI: [10.54708/2658-5014-SIIT-2023-no5-p88](https://doi.org/10.54708/2658-5014-SIIT-2023-no5-p88). EDN: ZWGBWG.
- Valeev R. S., Yusupov A. R. et al. "On the problem of the effective trade turnover supporting of a distribution company" // SIIT. 2024. V. 6, № 4(19), pp. 38–50. DOI: [10.54708/2658-5014-SIIT-2024-no4-p38](https://doi.org/10.54708/2658-5014-SIIT-2024-no4-p38). EDN: KSRBGK.
- Vokhmintcev A. V. "Methodology for solving the problem of simultaneous navigation and mapping based on a combination of visual and semantic characteristics of the environment" // SIIT. 2023. V. 5, № 3(12), pp. 136–155. DOI: [10.54708/2658-5014-SIIT-2023-no4-p136](https://doi.org/10.54708/2658-5014-SIIT-2023-no4-p136). EDN: EFOFFO.
- Gershteyn A. M., Terekhov A. N. "Bypassing Dangerous Sections of the Route as a Way to Improve Traffic Safety (Using St. Petersburg as an Example)" // Computer tools in education. 2023. No. 1, pp. 30–39. DOI: [10.32603/2071-2340-2023-1-30-39](https://doi.org/10.32603/2071-2340-2023-1-30-39). EDN: LBMCDH. (In Russian).
- Dzhuruk D.S. "Methodology for improving road safety on two-lane roads in areas with a high concentration of road accidents using the example of the Siberian Federal District". 2020. (In Russian). EDN: KCQGVV.
- Road accidents: coordinates, participants, and victims // Accident map; processed by "If to be precise", 2024. URL: [https://tochno.st/datasets/traffic\\_accidents](https://tochno.st/datasets/traffic_accidents). (In Russian).
- Intelligent Transport Systems: from 2024 to 2025. Online resource of the analytical agency TAdviser. [https://www.tadviser.ru/index.php/Статья:Интеллектуальные\\_транспортные\\_системы:\\_из\\_2024\\_в\\_2025?erid=2W5zFJaCHdr](https://www.tadviser.ru/index.php/Статья:Интеллектуальные_транспортные_системы:_из_2024_в_2025?erid=2W5zFJaCHdr) (In Russian).



- [Кап24] Капский Д. В., Богданович С. В., Куренков П. В., Филиппова Н. А. Вопросы с появлением транспортной отрасли в условиях развития подключённых транспортных средств // Интеллект. Инновации. Инвестиции. 2024. № 3. С. 64–73. DOI: [10.25198/2077-7175-2024-3-64](https://doi.org/10.25198/2077-7175-2024-3-64). EDN: [JIPRZJ](#).
- [Куз20] Кузьменко Е. А., Донченко Д. С., Рагозин В. О. Анализ данных для прогнозирования вероятности дорожно-транспортных происшествий с участием пешеходов // Инженерный вестник Дона. 2020. № 6. EDN: [ABLQZB](#).
- [Луц08] Луценко Е. В., Коржаков В. Е. Адаптивная семантическая информационная модель прогнозирования рисков совершения ДТП // Известия Алтайского государственного университета. 2008. № 4. EDN: [KYHERR](#).
- [Овч12] Овчаров Е. Б., Некрасов И. Б., Бакирова Э. И. Обзор используемых методов математического моделирования при разработке программного комплекса для оценки факторов и управления показателями безопасности дорожного движения // Управление экономическими системами: электронный научный журнал. 2012. № 9 (45). С. 10. EDN: [PWNKDL](#).
- [Ско25] Скоробогатченко Д. А., Кизим А. В. и др. Оценка влияния географических, погодных и социально-инфраструктурных факторов на объем потребления энергии на зарядных станциях для электротранспорта в условиях крупной городской агломерации // СИИТ. Т. 7, № 4 (23). С. 68–75. DOI: [10.54708/2658-5014-SIIT-2025-no4-p68](https://doi.org/10.54708/2658-5014-SIIT-2025-no4-p68). EDN: [LTFCNO](#).
- [Сай25] Сайтова Г. А., Габдуллина Э. Р. Методика определения проективного покрытия полей на основе дистанционного мониторинга // СИИТ. 2025. Т. 7, № 2 (21). С. 48–55. DOI: [10.54708/2658-5014-SIIT-2025-no2-p48](https://doi.org/10.54708/2658-5014-SIIT-2025-no2-p48). EDN: [XTKJHQ](#).
- Kapski, D. V., Bogdanovich, S. V., Kurenkov, P. V., Filippova, N. A. "Issues of transport sector improvement in the context of connected vehicles development" // Intellect. Innovations. Investments. 2024. № 3, pp. 64–73. DOI: [10.25198/2077-7175-2024-3-64](https://doi.org/10.25198/2077-7175-2024-3-64). EDN: [JIPRZJ](#). (In Russian).
- Kuzmenko E. A., Donchenko D. S., Ragozin V. O. "Data analysis for predicting probability of road accidents involving pedestrians" // Don Engineering Bulletin. 2020. No. 6. . EDN: [ABLQZB](#). (In Russian).
- Lutsenko E.V., Korzhakov V.E. "Adaptive semantic information model for predicting the risks of road accidents" // Bulletin of the Altai State University. 2008. № 4. EDN: [KYHERR](#). (In Russian).
- Ovcharov E.B., Nekrasov I.B., Bakirova E.I. "Overview of methods for mathematical modeling employed in developing software for assessing factors and managing road traffic safety indicators" // Management of economic systems. 2012. № 9(45). С. 10. EDN: [PWNKDL](#). (In Russian).
- Skorobogatchenko D. A., Kizim A. V. et al. "Assessment of the impact of geographical, weather, and socio-infrastructure factors on energy consumption at electric vehicle charging stations in a large urban agglomeration" // SIIT. 2025. V. 7, № 4(23), pp. 68–75. DOI: [10.54708/2658-5014-SIIT-2025-no4-p68](https://doi.org/10.54708/2658-5014-SIIT-2025-no4-p68). EDN: [LTFCNO](#).
- Saitova G. A., Gabdullina E. R. "Method for determining projective field coverage based on remote monitoring" // SIIT. 2025. V. 7, № 2(21), pp. 48–55. DOI: [10.54708/2658-5014-SIIT-2025-no2-p48](https://doi.org/10.54708/2658-5014-SIIT-2025-no2-p48). EDN: [XTKJHQ](#).

## ОБ АВТОРАХ | ABOUT THE AUTHORS

### ШАЙМАРДАНОВ Айгиз Фанузович

Уфимский университет науки и технологий, Россия.  
[aigiz2001@list.ru](mailto:aigiz2001@list.ru).  
 Магистр по направлению «Программная инженерия».

### САЗОНОВА Екатерина Юрьевна

Уфимский университет науки и технологий, Россия.  
[sazonova.eyu@ugatu.ru](mailto:sazonova.eyu@ugatu.ru) ORCID [0000-0001-8834-2635](https://orcid.org/0000-0001-8834-2635).  
 Доц. каф. вычислительной математики и кибернетики. Дипл. экон.-математик (Уфимск. гос. авиац. техн. ун-т, 2011). Канд. техн. наук по системному анализу, управлению и обработке информации (Там же, 2015). Иссл. в обл. управления, сист. анализа, интел. обр. данных.

### СМЕТАНИНА Ольга Николаевна

Уфимский университет науки и технологий, Россия.  
[smetanina.on@ugatu.ru](mailto:smetanina.on@ugatu.ru) ORCID [0000-0001-6970-1110](https://orcid.org/0000-0001-6970-1110).  
 Проф. каф. вычислительной математики и кибернетики. Дипл. инж.-электрик (Уфимск. авиац. ин-т, 1985). Д-р техн. наук по управлению в соц. и экон. системах (Уфимск. гос. авиац. техн. ун-т, 2012). Иссл. в обл. интелл. поддержки принятия решений.

### SHAYMARDANOV Aigiz Fanuzovich

Ufa University of Science and Technology, Russia.  
[aigiz2001@list.ru](mailto:aigiz2001@list.ru).  
 Master's degree in "Software Engineering".

### SAZONOVA Ekaterina Yurevna

Ufa University of Science and Technology, Russia.  
[sazonova.eyu@ugatu.ru](mailto:sazonova.eyu@ugatu.ru) ORCID [0000-0001-8834-2635](https://orcid.org/0000-0001-8834-2635).  
 Assoc. Prof. of Computational Mathematics and Cybernetics dept. Dipl. Economist-Mathematician (Ufa State Aviation Tech. Univ., 2011). Cand. Tech. Sci. on System Analysis, Management and Information Processing (ibid., 2015). Research in the field of management, systems analysis, data mining.

### SMETANINA Olga Nikolaevna

Ufa University of Science and Technology, Russia.  
[smetanina.on@ugatu.ru](mailto:smetanina.on@ugatu.ru) ORCID [0000-0001-6970-1110](https://orcid.org/0000-0001-6970-1110).  
 Prof. Computational Mathematics and Cybernetics Dept. Dipl. Eng.-Electrician (Ufa Aviat. Institut, 1985). Dr. Technical Sciences in social and economic management systems (Ufa State Aviat. Tech. University, 2012). Research in the field of intelligent decision support.

## МЕТАДАННЫЕ | METADATA

**Заглавие:** Построение маршрутов с учетом спрогнозированного риска дорожно-транспортных происшествий на основе исторических данных.

**Авторы:** Шаймарданов А. Ф., Сазонова Е. Ю., Сметанина О. Н.

**Title:** Construction of routes taking into account the predicted risk of road accidents based on historical data.

**Authors:** Shaymardanov A. F., Sazonova E. Yu., Smetanina O. N.

**Аннотация:** Увеличение количества транспортных средств и водителей на дорогах ведет к росту количества дорожно-транспортных происшествий (ДТП). Одним из способов улучшения данной ситуации является внедрение интеллектуальных систем обработки транспортной информации, которые позволят эффективно управлять дорожным движением, выявлять потенциально опасные участки дорог, выбирать маршруты с учётом риска возникновения ДТП. В статье предлагается разработать решение поиска рациональных маршрутов с учетом оценки риска ДТП на основе исторических данных. Авторами приведены постановка задачи построения рациональных маршрутов, результаты анализа существующих программных решений в области построения маршрутов с учетом рискованной составляющей на дорогах. Анализ позволил сделать вывод, что существующие решения предоставляют базовый функционал для анализа дорожной обстановки, но не предлагают детального анализа сегментов маршрута с учетом риска ДТП; большинство систем фокусируются на текущих данных о пробках и дорожных событиях, игнорируя долгосрочный анализ и прогнозирование. В работе решаются две задачи – прогнозирование риска ДТП и построение рационального маршрута с учетом выявленного риска. Для решения задачи прогнозирования риска была выбрана библиотека CatBoost, разработанная Яндексом. CatBoost отличается высокой эффективностью при работе с категориальными признаками и обладает встроенными средствами предотвращения переобучения. Для обучения модели машинного обучения был использован набор данных «Дорожно-транспортные происшествия: координаты, участники и пострадавшие». Для построения безопасного маршрута и визуализации его на карте используются OpenStreetMap в сочетании с Open Source Routing Machine. Предложенные решения имеют практическую ценность и могут служить основой для дальнейших исследований и развития интеллектуальных систем анализа транспортной информации.

**Ключевые слова:** Построение маршрутов; прогнозирование риска ДТП; градиентный бустинг; CatBoost.

**Язык:** Русский.

Статья поступила в редакцию 15 июля 2025 г.

**Abstract:** The increasing number of vehicles and drivers on the roads leads to an increase in the number of road accidents. One way to improve this situation is to implement intelligent transport information processing systems that will enable effective traffic management, identify potentially dangerous road sections, and select routes taking into account the risk of accidents. This article proposes developing a solution for finding rational routes that takes into account accident risk assessment based on historical data. The authors present the problem statement for constructing rational routes and the results of an analysis of existing software solutions for constructing routes that take into account road risk. The analysis revealed that existing solutions provide basic functionality for analyzing traffic conditions but do not offer a detailed analysis of route segments taking into account accident risk. Most systems focus on current data on traffic jams and road events, ignoring long-term analysis and forecasting. This paper addresses two problems: predicting the risk of accidents and constructing a rational route taking into account the identified risk. The CatBoost library developed by Yandex was selected for the risk prediction task. CatBoost is highly effective when working with categorical features and has built-in overfitting prevention. The "Road Accidents: Coordinates, Participants, and Victims" dataset was used to train the machine learning model. OpenStreetMap in combination with the Open Source Routing Machine is used to plot a safe route and visualize it on a map. The proposed solutions have practical value and can serve as a basis for further research and development of intelligent systems for analyzing transport information.

**Key words:** Route planning; accident risk prediction; gradient boosting; CatBoost.

**Language:** Russian.

The article was received by the editors on 15 July 2025.