

Автономная некоммерческая образовательная организация
высшего образования

«НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»

Научный центр информационных технологий и искусственного интеллекта
направление «Математическое моделирование в биомедицине и геофизике»

К ЗАЩИТЕ ДОПУСТИТЬ

Научный руководитель направления

«Математическое моделирование в
биомедицине и геофизике»

член-корр. РАН, д.ф-м.н., профессор

_____ / Ю.В. Василевский

« ____ » _____ 2024 г.

**РАЗРАБОТКА СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ЛЕСНЫХ ПОЖАРОВ
С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ**

Магистерская диссертация

по направлению подготовки 01.04.02 Прикладная математика и информатика
(направленность (профиль) «Математическое моделирование в биомедицине и
нефтегазовом инжиниринге»)

Студент гр. М01ММ-22

_____ / Е.С. Удалова

« ____ » _____ 2024 г.

Научный руководитель магистерской
диссертации

Научный руководитель направления
«Математическое моделирование в
биомедицине и геофизике» Научного
центра информационных технологий и
искусственного интеллекта, д.ф-м.н.

_____ / Ю.В. Василевский

« ____ » _____ 2024 г.

Федеральная территория «Сириус», 2024

Автономная некоммерческая образовательная организация
высшего образования
«НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»
Научный центр информационных технологий и искусственного интеллекта
Направление «Математическое моделирование в биомедицине и нефтегазовом
инжиниринге»

УТВЕРДИТЬ

Научный руководитель направления
«Математическое моделирование в
биомедицине и геофизике»

член-корр. РАН, д.ф.-м.н., профессор
ученая степень, звание

_____ Ю.В. Василевский
« ___ » _____ 2024 г.

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

на выполнение выпускной квалификационной работы

обучающегося по направлению подготовки 01.04.02 Прикладная математика и
информатика

направленность (профиль) «Математическое моделирование в биомедицине и
нефтегазовом инжиниринге»

Удаловой Елены Сергеевны

1. Тема: «Разработка системы прогнозирования лесных пожаров с помощью методов машинного обучения»
2. Цель: на основе анализа факторов, влияющих на пожароопасность, разработать систему, прогнозирующую вероятность возникновения лесного пожара с помощью методов машинного обучения, и провести апробацию эффективности
3. Задачи:
 - Провести обзор текущего состояния вопроса путем анализа факторов, влияющих на вероятность возникновения пожаров, и существующих подходов к прогнозированию природных пожаров и явлений пожарной опасности.
 - Произвести валидацию наиболее перспективного решения на основе машинного обучения для проверки корректности работы на территории России и выявить недостатки модели.
 - Разработать новую модель с учетом выявленных точек роста для одного из регионов России.
4. Рабочий график (план) выполнения выпускной квалификационной работы:

№	Перечень заданий	Сроки выполнения
1	Поиск и анализ методов прогнозирования пожарной опасности	30.01.2024 – 15.02.2024
1	Поиск и анализ источников данных о реальных пожарах	15.02.2024 – 28.02.2024
2	Тестирование работы модели на данных о реальных пожарах	1.03.2024 – 31.03.2024
2	Анализ факторов, влияющих на развитие и распространение пожара	01.04.2024 - 30.04.2024
3	Доработка модели машинного обучения	01.05.2024 - 23.05.2024

Дата выдачи: «__» _____ 2024

Руководитель:

_____ Ю.В. Василевский
подпись И.О. Фамилия

Задание принял к исполнению:

Студент группы М01ММ-22

_____ Е.С. Удалова
подпись И.О. Фамилия

«__» _____ 2024

Реферат

Выпускная квалификационная работа, 66 страниц, 18 рисунков, 9 таблиц, 63 источника.

ПОЖАРНАЯ ОПАСНОСТЬ, ЛЕСНЫЕ ПОЖАРЫ, ГРАДИЕНТНЫЙ БУСТИНГ, МАШИННОЕ ОБУЧЕНИЕ, ПРОСТРАНСТВЕННАЯ ИНТЕРПОЛЯЦИЯ, ВРЕМЕННЫЕ РЯДЫ

Объектом исследования являются методы прогнозирования лесной пожарной опасности, а предметом – алгоритм машинного обучения для ее прогнозирования.

Цель работы – на основе анализа факторов, влияющих на пожароопасность, разработать систему, прогнозирующую вероятность возникновения лесного пожара на географической сетке с помощью методов машинного обучения, и провести апробацию эффективности.

В данной работе были выявлены факторы, влияющие на лесную пожароопасность, на основе которых была обучена модель градиентного бустинга, прогнозирующая вероятность возникновения лесного пожара. В ее основу легли погодные условия, антропогенные и геоморфологические факторы территории, а также пространственное распределение опасных зон возникновения возгораний, которое было составлено интерполяцией данных о классах пожарной опасности точек возникновения реальных пожаров, произошедших за период, предшествующий прогнозному. Полученная в работе модель показала результат лучше, чем модель, выбранная в качестве базовой, в задаче прогнозирования природных возгораний в период высокой пожароопасности.

Abstract

Graduation Thesis, 66 pages, 18 figures, 9 tables, 63 sources.

FIRE HAZARD, FOREST FIRES, GRADIENT BOOSTING, MACHINE LEARNING, SPATIAL INTERPOLATION, TIME SERIES

The object of the study is methods for predicting forest fire danger, and the subject is a machine learning algorithm for predicting it.

The objective of the research is to develop a system based on the analysis of factors affecting fire hazard, predicting the likelihood of a forest fire using machine learning methods, and to test the effectiveness.

In this research, factors affecting forest fire hazard were identified, on the basis of which a gradient boosting model was trained, predicting the probability of a forest fire. It is based on weather conditions, anthropogenic and geomorphological factors of the territory, as well as the spatial distribution of hazardous fire zones, which was compiled by interpolating data on fire hazard classes of points of occurrence of real fires that occurred during the period preceding the forecast one. The model obtained in the work showed a better result than the model chosen as the basic one in the task of predicting natural fires during a period of high fire danger.

Оглавление

Введение.....	7
1 ТЕОРЕТИЧЕСКИЙ АНАЛИЗ ПОДХОДОВ К ПРОГНОЗИРОВАНИЮ ПРИРОДНЫХ ПОЖАРООПАСНЫХ СИТУАЦИЙ.....	10
1.1 Природные пожары.....	10
1.2 Факторы возникновения природных пожаров.....	14
1.3 Методики моделирования и прогнозирования пожарной опасности.....	19
1.3.1 Эмпирические подходы к определению пожарной опасности	20
1.3.2 Аналитические методы моделирования пожарной опасности	24
1.3.3 Методы прогнозирования пожарной опасности на основе алгоритмов машинного и глубокого обучения.....	26
2 ВАЛИДАЦИЯ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ПОЖАРООПАСНОСТИ И ОПРЕДЕЛЕНИЕ КРИТЕРИЕВ, ПОВЫШАЮЩИХ ЭФФЕКТИВНОСТЬ ЕЕ РАБОТЫ.....	29
2.1 Анализ принципов работы модели.....	29
2.1.1 Методы градиентного бустинга	30
2.1.2 Сбор данных	31
2.1.3 Предобработка данных.....	32
2.1.4 Обучение модели	34
2.2 Валидация работы модели на данных о реальных пожарах Российской Федерации.....	37
2.2.1 Изучение реальных данных	37
2.2.2 Сравнение прогнозных значений с реальными данными.....	39
2.2.3 Поиск доверительного порога	41
3 УЛУЧШЕНИЕ РЕШЕНИЯ И ТЕСТИРОВАНИЕ ГИПОТЕЗ НА ПРИМЕРЕ ОДНОГО ИЗ РЕГИОНОВ РОССИИ.....	46
3.1 Построение карт пожарной опасности	47
3.2 Разработка новой модели	53
Заключение	58
Список использованных источников	59

Введение

Пожары являются одними из самых разрушительных и часто встречающихся стихийных бедствий, которые могут привести к серьезным последствиям, в том числе человеческим жертвам, экономическому и экологическому ущербу. Среди всех типов пожаров лесные представляют собой отдельную значительную проблему из-за своей масштабности, сложности в регулировании и потенциальной разрушительности.

Средняя продолжительность сезона лесных пожаров за последние десятилетия увеличилась на 20% из-за активного землепользования и глобального потепления [1], что привело к повышению температуры, более частым засухам и снижению влажности воздуха [2]. Территориальный масштаб воздействия лесных пожаров также увеличился, а их частота и интенсивность растут во всем мире [3].

Леса, покрывающие примерно 45.6% территории России [4], ежегодно оказываются под угрозой разрушительных пожаров. Так, каждый год происходит не менее 30 тысяч лесных возгораний, общий ущерб от которых достигает 20 миллиардов рублей. Причем, убытки включают в себя следующие расходы:

- Стоимость потери древесины;
- Стоимость сгоревших объектов;
- Расходы на тушение лесных пожаров;
- Другие потери.

Важная часть бюджета, закладываемая при его планировании, выделяется на материально-техническое обеспечение: использование специализированной техники и ресурсов, оплату труда сотрудников служб МЧС, переброску техники. Так, в 2023 году было произведено более 250 перебросок сил и средств в рамках межрегионального маневрирования – значение превышает прошлогоднее в полтора раза [5].

Несмотря на то, что полностью устранить риск возникновения природных пожаров невозможно, необходимо стремиться к регулированию и уменьшению влияния этого природного явления. Один из способов – предиктивная аналитика, выявляющая потенциальные «горячие зоны», возникновение возгорания в которых наиболее вероятно. Подобные решения совместно с моделированием распространения огня позволят планировать бюджеты и нагрузку на работников пожарной охраны.

На данный момент аналитика производится путем расчёта классов пожарной опасности регионов и районов. Для служб лесопожарной охраны используется регламент с классификацией от 2011 года, в основе которой лежит методика 1949 года, что свидетельствует о нечастом обновлении подобных алгоритмов и методов. Несмотря на быстрые темпы роста технологических возможностей и широкую доступность пространственных данных, подобные технологии недостаточно развиты в данном секторе. Также стоит отметить сложность и комплексность условий возникновения пожара, зависящих от многочисленных факторов, мониторинг которых ресурсозатратен при ручной обработке.

Цель магистерской диссертации – на основе анализа факторов, влияющих на пожароопасность, разработать систему, прогнозирующую вероятность возникновения лесного пожара с помощью методов машинного обучения, и провести апробацию эффективности.

Для достижения заявленной цели были выделены следующие задачи:

1. Провести обзор текущего состояния проблемы путем анализа факторов, влияющих на вероятность возникновения пожаров, и существующих подходов к прогнозированию природных пожаров и явлений пожарной опасности.

2. Произвести валидацию наиболее перспективного решения на основе машинного обучения для проверки корректности работы на территории России и выявить недостатки модели.

3. Разработать новую модель с учетом выявленных точек роста для одного из регионов России.

1 ТЕОРЕТИЧЕСКИЙ АНАЛИЗ ПОДХОДОВ К ПРОГНОЗИРОВАНИЮ ПРИРОДНЫХ ПОЖАРООПАСНЫХ СИТУАЦИЙ

В рамках исследования в данной главе будет проведен анализ ключевых аспектов прогнозирования природных пожаров, в том числе:

- Обзор статистики природных пожаров, их видов и характеристик,
- Определение перечня факторов, влияющих на возникновение природных возгораний,
- Анализ имеющихся систем мониторинга и прогнозирования лесной пожарной опасности.

Таким образом, систематизация этих аспектов позволит не только углубить понимание причин возникновения и динамики развития природных пожаров, но и разработать рекомендации по улучшению систем мониторинга и прогнозирования, что является важным шагом на пути к минимизации ущерба от пожаров и обеспечению безопасности населения и экосистем страны.

1.1 Природные пожары

Природные пожары представляют собой сложное и многогранное явление, которое охватывает различные формы возгораний в экосистемах. Природные пожары играют важную роль в биоценозах, участвуя в цикле восстановления и обновления компонентов, однако их неконтролируемое распространение может привести к значительным экологическим, экономическим и социальным последствиям.

Такие пожары имеют несколько классификаций в зависимости от места возникновения: лесные, степные, торфяные, подземные. Рассмотрим их подробнее.

Площади пройденных пожарами территорий – это ключевой индикатор для оценки масштабов и последствий в различных экосистемах. На рисунке 1.1 представлен график, отображающий площади, пройденные различными

типами пожаров за период с 2002 по 2022 год в России. Он построен с использованием библиотеки *Matplotlib* для языка программирования Python на основе данных из открытого датасета [6], сформированного на основе информации из Глобальной информационной системы о лесных пожарах (The Global Wildfire Information System) по спутниковым изображениям. Графики отражают динамику изменения площадей, которые были охвачены пожарами в разные годы. Торфяные и подземные пожары не представлены в данной статистике, а лесные пожары в общем понимании разделены на возгорания на непосредственно лесных территориях и на площадях, покрытых кустарниками и лугами.

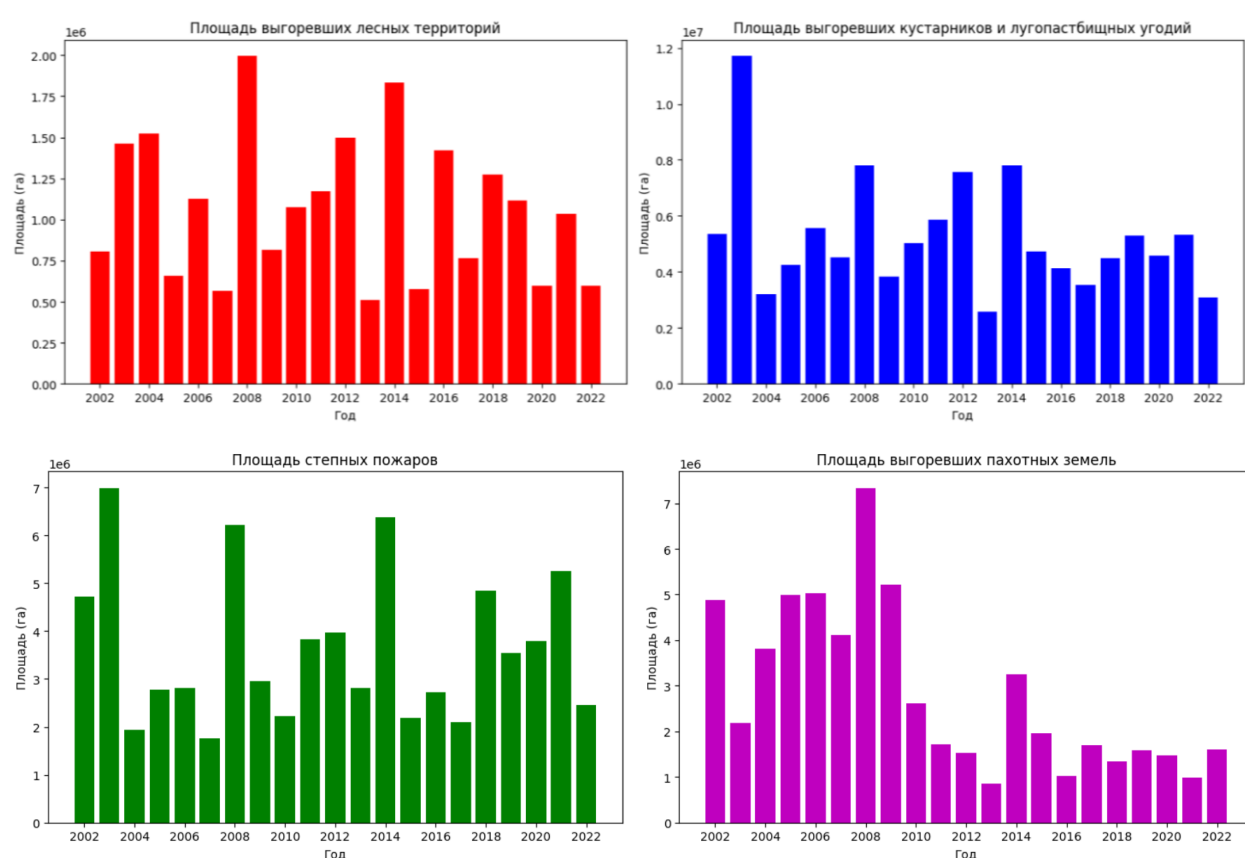


Рисунок 1.1 – Площади земель различных категорий, сгоревших в результате пожаров в России за 2002-2022 гг.

Так, лесные пожары – это наиболее распространенная форма природных пожаров в России. Возгорания охватывают разные уровни леса – от верхушек и низов, где в том числе располагается подлесок, кустарники и травянистая

растительность, до слоёв почвы. На рисунке 1.2 отображены данные о местах возникновения и типах природных пожаров за 2012-2021 гг. в России [7] с помощью библиотеки *GeoPandas*. Можно заметить, что очаги лесных пожаров покрывают всю территорию страны, однако более выражены на Урале, в Сибири и на Дальнем Востоке. В Европейской части, на Арктической зоне и на Камчатке они встречаются и фиксируются значительно реже. Это может быть связано как с хорошими противопожарными мерами, так и с условиями, неблагоприятными для развития возгорания.

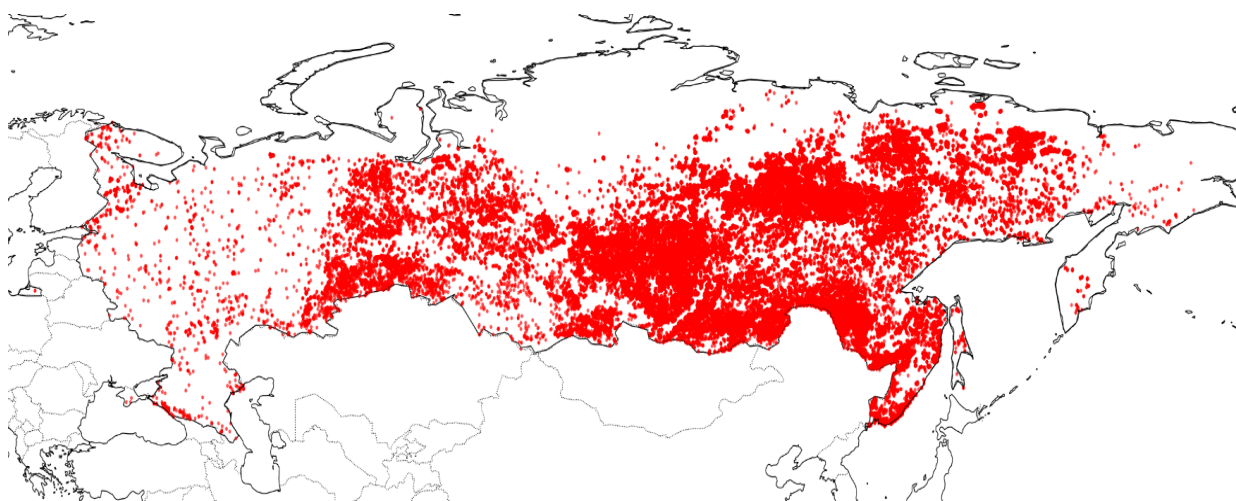


Рисунок 1.2 – Карта лесных пожаров, произошедших в России за 2012-2021 гг.

Также важно рассмотреть другие виды природных пожаров, поскольку зачастую контролируемый и неконтролируемый пал, возникающий из-за выжигания травы под надзором противопожарной службы или без него, под действием ветра распространяется на лесные насаждения. Сжигание – экономичный способ очистки полей от прошлогодней сухой растительности, однако в случае несоблюдения техники безопасности ситуация может выходить из-под контроля. На рисунке 1.3 представлена карта расположения палов травы за 2012-2021 годах в России.

Контролируемый пал распределен по всей территории страны, поскольку мероприятия по выжиганию травы проводятся во многих регионах.

Локализация палов в восточной части России пересекается с расположением лесных пожаров, что говорит о возможных переходах одного пожара в другой.

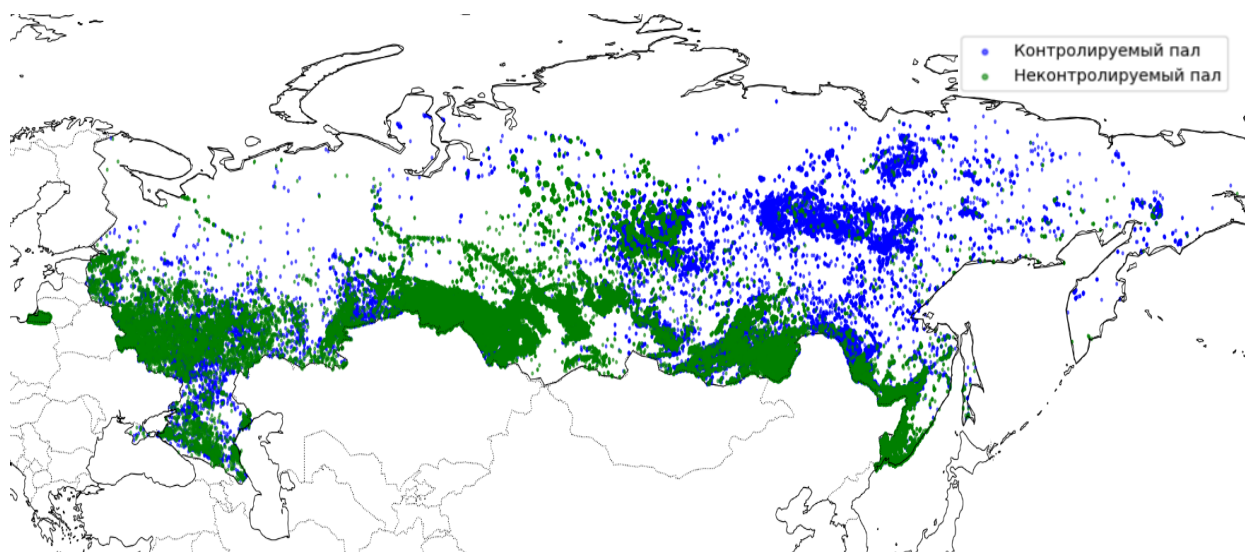


Рисунок 1.3 – Карта палов травы, произошедших в России за 2012-2021 гг.

Торфяные пожары локализованы в болотистых регионах, где могут тлеть в почве длительное время и выделять значительное количество углекислого газа и других вредных веществ. Они трудно поддаются тушению из-за глубокого залегания очагов возгорания. По этой же причине они хуже определяются по снимкам со спутников и мало представлены в открытых наборах данных по прошедшим в России природным пожарам (рисунок 1.4).



Рисунок 1.4 – Карта торфяных пожаров, произошедших в России за 2012-2021 гг.

Таким образом, можно прийти к выводу о том, что природные пожары неравномерно распределены по территории России, поэтому важно перейти к рассмотрению факторов, способствующих их возникновению. Они обусловлены множеством условий, которые варьируются в зависимости от регионов. Их понимание позволит разработать более точные инструменты предиктивной противопожарной аналитики.

1.2 Факторы возникновения природных пожаров

Согласно определению Н.П. Курбатского [8], пожарная опасность – вероятность возникновения природного пожара, которую можно количественно оценить как вероятность возникновения события (формула 1.1) и качественно описать пирологическими характеристиками потенциального очага возгорания и окружающих его условий:

$$P = f(F, t), \quad (1.1)$$

где $F = (f_1; f_2 \dots f_n)$ – факторы, влияющие на пожарную опасность,

n – число значимых факторов,

t – горизонт прогнозирования.

Понимание факторов, влияющих на возникновение и распространение пожара, имеет важное значение для их прогнозирования и предотвращения.

Для возникновения пожара необходимы компоненты, называемые «Треугольником огня»:

- Горючее вещество, в том числе любой легковоспламеняющийся и сгораемый материал;
- Окислитель, которым чаще всего выступает кислород;
- Источник тепла.

По статистике авиационной охраны лесов России [9], причиной большинства пожаров, то есть источником, становятся действия человека. Это может быть как неосторожное обращение с огнем, так и последствия охотничьего промысла или неправильная утилизации отходов – например,

воспламенение лесной подстилки из-за фокусировки солнечных лучей стеклянным осколком. В данном вопросе важно учитывать, что подобные антропогенные возгорания происходят недалеко от жилых зон – несмотря на низкую возможность прогнозирования, они быстро обнаруживаются и эффективно тушатся, не достигая значительных площадей тления.

Источниками тепла в случае естественных лесных пожаров зачастую становятся сильный солнечный свет, удар молнии или процесс гниения торфяников. Удаленные от мест обитания людей возгорания, возникающие по природным причинам, обнаруживать и тушить затруднительно, поэтому фронт огня начинает стремительно распространяться, бесконтрольно охватывая гектары территорий.

Интенсивность солнечной радиации связывают с характеристиками лесных пожаров. Исследованию влияния климатических условий на лесные пожары посвящены работы многих исследователей, в том числе Г.Н. Коровина [10], С.Г. Аксёнова [11], М.И. Петрова [12], И.В. Латышева [13]. В международной научной литературе также активно изучается эта зависимость: влияние изменения климата на пожароопасность, в том числе в лесах северо-западных регионов Тихого океана [14], в Калифорнии [15] и других территориях, подверженных риску лесных пожаров. Помимо этого, климатические изменения считаются ключевым фактором, вызывающим пожары на глобальном уровне [16, 17].

Грозная пожароопасность, по мнению П.С. Бурлакова [18], зачастую связана с аномалиями магнитного поля и неоднородностью геологического строения территорий, на которых образуются сухие грозы – природные явления, при которых электрические разряды накапливаются в облаках без осадков. Так происходит, поскольку капли дождя испаряются, не достигнув поверхности. Это довольно распространенная причина природных возгораний – в июле 2016 года на сухие грозы приходилось до 62% пожаров в России.

Дополнительными источниками тепла также могут становиться дымовые эмиссии – выбросы CO₂ от соседних пожаров, которые задерживают теплый воздух, предотвращая излучение солнечного тепла обратно в космос.

Распространение лесных пожаров после их возникновения определяется горючестью материала, зависящей от ряда факторов. Важнейшими из них являются содержание смол и масел, а также гигроскопичность – способность растений аккумулировать влагу. Состав и количество смол и масел зависит от разновидности древесины и условий, в которых растут деревья [19]. Большое количество накопленного горючего материала, по мнению Ю.Н. Самсонова и Г.А. Иванова [20], влияет на интенсивность горения. Однако, как они утверждают, это не является определяющим фактором, поскольку важную роль играют погодные условия, формирующие пирологические свойства растительности. Такими условиями, например, являются в том числе продолжительность засушливого сезона и заморозки: ночные отрицательные температуры провоцируют процесс вымораживания, при котором влажность воздуха в дневные часы резко падает. Помимо прочего, в периоды без осадков не происходит самогашения уже существующих пожаров.

Точка росы является важным метеорологическим показателем, который указывает на температуру, при которой воздух становится насыщенным влагой и начинается конденсация. Низкая точка росы указывает на сухие условия, что способствует высыханию растительности и увеличивает риск возникновения и распространения лесных пожаров. В статье Е.В. Болдановой [21], посвященной прогнозированию количества и площадей лесных пожаров, рассматриваются различные методы расчета накопленного дефицита точки росы. Наиболее сильная корреляционная зависимость между числом лесных пожаров и накопленным дефицитом точки росы была выявлена на сроках свыше 10 дней.

Кроме того, Е.А. Лупян и С.А. Барталев в научной работе [22], посвященной анализу лесных пожаров за 2001-2016 года, отметили важность учета пожароопасности сезона. Амплитудные пики пожарной опасности на

несколько недель выделяются весной и летом. Действительно, данный тренд сохраняется и в следующем десятилетии, что иллюстрирует рисунок 1.5. На графике заметно выражена сезонность пожаров: пики приходятся на весенние (апрель) и летние месяцы (июль-август). В сентябре и осенних месяцах наблюдается уменьшение числа пожаров, хотя в некоторой литературе выделяют также и менее заметный осенний пик.

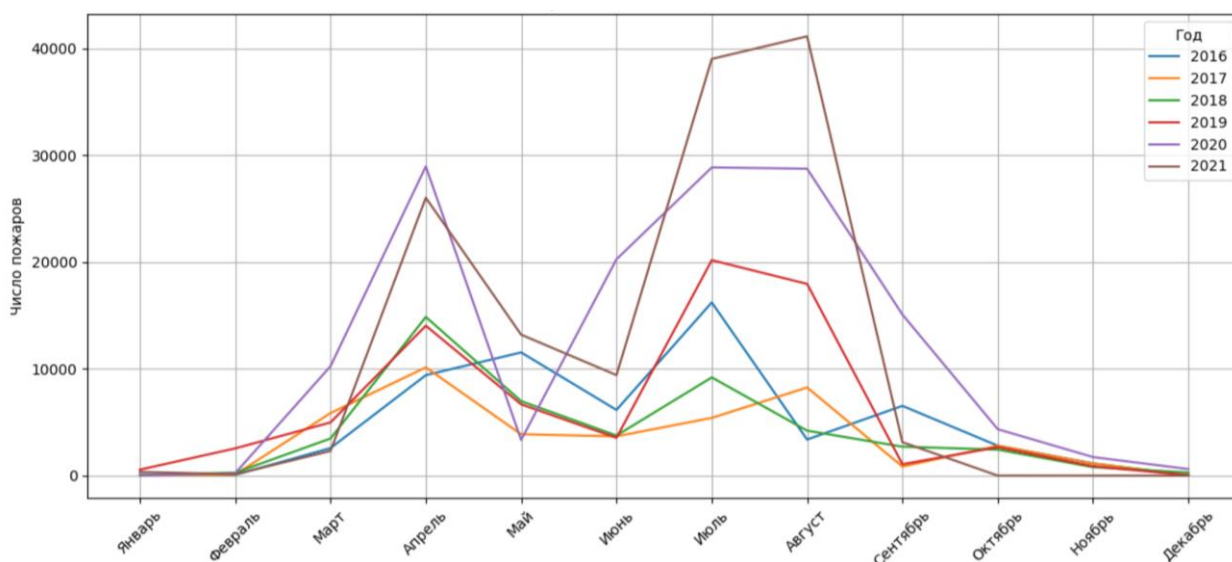


Рисунок 1.5 – Число пожаров по месяцам с 2016 по 2021 годы

В исследованиях также описывается неравномерность распределения числа точек возникновения пожаров в зависимости от типа территории и растительности. Основная масса пожаров (до 41%) приходится на лиственные леса, а самая маленькая доля происходит в тундре, болотах и хвойных кустарниках. Более того, для разных типов леса наблюдаются отличающиеся графики распределения числа точек пожаров по сезонам: лиственные леса начинают гореть раньше, а у хвойных лесов наблюдаются ярко выраженные пики практически одинаковой амплитуды и весной, и летом, и осенью.

Тип древесины играет более значимую роль в возникновении пожаров из-за различных особенностей строения. К таким факторам можно отнести глубину расположения корневой системы, которая влияет на вероятность падения дерева в случае низовых пожаров. Также имеет значение сомкнутость кроны, определяя возможность распространения верхового пожара, который

перекидывается от дерева к дереву по верхам. Так, например, в Сибири в основном происходят низовые пожары, поскольку для верховых нужна сомкнутость крон деревьев или ярко выраженная ярусность растительности, через которую огонь будет подниматься до верхушек.

Значение имеет состояние и статус эксплуатации леса. Наиболее пожароопасными считаются территории с вырубленными деревьями, поскольку отходы лесозаготовок легко воспламеняемы [23]. Также опасны пораженные болезнями участки леса и территории, на которых ведутся сельскохозяйственные работы – около 3% лесных пожаров возникает из-за сельскохозяйственных палов [22].

Тип подстилающей местности также влияет на вероятность возникновения возгорания: по статистике, «до 44% лесных пожаров зарождаются на территориях, покрытой травянистой растительностью» [24, с. 168]. Исследования [25, 26] показывают, что количество энергии для воспламенения напрямую зависит от структуры материала, плотности и исходного содержания влаги в растениях, а также имеются значительные различия в воспламеняемости травянистых видов растений.

К факторам, влияющим не только на возникновение, но и распространение пожаров, относятся скорость и направление ветра – они способствуют масштабированию площади, охваченной огнем, увеличивая его интенсивность и температуру, а также меняя направление распространения пожара. Современные исследования [27] в данной области подтверждают значимость этих факторов.

Таким образом, была составлена следующая классификация факторов:

- Погодные факторы:

Температура: дневная температура, ночная температура, интенсивность солнечной радиации;

Ветер: скорость ветра, направление, наличие дымовых эмиссий;

Осадки: число дней без осадков, величина осадков, влажность, точка росы, активность молний;

- Антропогенные факторы: близость к населенным пунктам, дорогам (разных типов), населенность, статус эксплуатации местности;

- Геоморфологические факторы: близость к источникам воды, тип и состояние растительности и подстилающей почвы, глубина корневой системы, ярусность, мощность отмершей растительности;

- Прочие факторы, в том числе периодические: месяц года, день недели.

Полученная классификация факторов, определяющих опасность возникновения пожаров, станет основой для дальнейшей оценки пожароопасности. Ее применение позволит учитывать разнообразные взаимосвязи и получить комплексную характеристику пожарной опасности.

1.3 Методики моделирования и прогнозирования пожарной опасности

В области прогнозирования лесных пожаров существует множество решений, направленных на предоставление различных видов информации и применения широкого спектра подходов и технологий.

Классическая задача для реагирования на пожары – детектирование активного горения. При организации мониторинга природных пожаров часто используется информация с низкоорбитальных спутниковых систем, например, данных со спектрорадиометров MODIS (Moderate Resolution Imaging Spectroradiometer), получаемых со спутников Terra и Aqua, а также спектрорадиометра VIIRS (Visible Infrared Imaging Radiometer Suite) со спутника NOAA-21. Существуют алгоритмы для автоматизированного выделения «горячих точек» на инфракрасных каналах спутниковых изображениях, соответствующих участкам с высокотемпературными объектами, к которым, помимо антропогенных объектов и тепловых аномалий Земли, также относятся и природные пожары. Такие температурные аномалии называются термоточками – зарегистрированными в момент пролета спутника значительными повышениями температуры на поверхности земли по сравнению с соседними участками. На данный момент развит ряд

автоматизированных подходов и технологий, позволяющих не только детектировать места возгораний на отдельных сеансах спутникового наблюдения, но и анализировать их характеристики развития.

Таким образом, в настоящее время возможно регулярно и оперативно получать достоверный набор термоточек, соответствующих участкам горения пожаров.

Помимо детекции, другая важная задача – прогнозирование лесных пожаров с различными объектами исследований. К основным можно отнести следующие направления: определение вероятности и локации возникновения лесного пожара, оценка возможного распространения огня и потенциальных последствий.

Моделирование угрозы возникновения пожаров включает в себя последовательную оценку и прогнозирование риска возгорания. Оценка включает в себя разделение территории на участки с присваиванием каждой образовавшейся ячейке сетки показателя пожарной опасности на текущий момент времени. Прогнозирование предполагает предсказание вероятности развития ситуации со временем, основываясь на предыдущих данных и знаниях о том, какие факторы могут измениться.

В научных трудах выделяют следующие типы моделей лесных пожаров: теоретические (математические) и эмпирические (статистические).

1.3.1 Эмпирические подходы к определению пожарной опасности

Статистические методы прогнозирования пожарной опасности играют немаловажную роль в управлении рисками и обеспечении безопасности в лесных и природных зонах. Основываясь на исторических данных о пожарах, климатических показателях и других значимых факторах, эти методы позволяют выявлять закономерности и тенденции пожароопасности. В таких моделях агрегируются и анализируются данные о скорости распространения лесного пожара при различных вариациях ряда параметров, а также

определяются коэффициенты корреляции по каждой независимой переменной. Этот подход не описывает механизм явления, и полученные зависимости нельзя применить за пределами области использованных статистических данных. Тем не менее, такие модели могут давать прогнозы с определенной степенью вероятности.

Базово задача прогнозирования вероятности возникновения пожарной опасности может быть решена путем распределения территорий по классам пожарной опасности и создания карт растительных горючих материалов посредством выделения классов растительного покрова, на которых природные пожары распространяются схожим образом. Институтом леса им. В.Н. Сукачева разработаны методы оценки растительных горючих материалов по спутниковым изображениям высокого и сверхвысокого разрешения [28, 29]. Подобная оценка может достигаться различными подходами. Один из них – выделение типичных пирологических характеристик растительности, которые можно использовать для приблизительного описания свойств определенного участка.

Так, леса России были классифицированы на пять типов природной пожарной опасности и сформированы в шкалу, которая до сих пор применяется для противопожарного устройства лесов с небольшими изменениями. Однако из-за необходимости отнесения территории лишь к одному классу, несмотря на растительное многообразие участка, концепция обеспечивает лишь общее, приближенное представление о растительных горючих материалах в данной области. В связи с этим такой подход может дать лишь упрощенное представление, не охватывающее реальную сложность условий, особенно для масштабного картографирования. Подобная классификация дает представление о горимости участков без учета реальных метеорологических условий и комплексности проблемы, опираясь в основном на данные о составе лесов [30].

Оценка пожарной опасности в лесах с учетом погодных условий основывается на расчете комплексного показателя, разработанного В.Г.

Нестеровым и учитывающего совокупность факторов, определяющих влажность лесных горючих материалов. Комплексный показатель определяется по формуле 1.2:

$$\text{КП} = \sum_{n=1}^i t^i \cdot (t^i - r^i), \quad (1.2)$$

где n – количество дней, прошедших с даты последнего выпадения осадков более 3 мм;

t^i – температура воздуха по показаниям сухого термометра в i -й день;

r^i – точка росы для температуры воздуха, то есть разность показаний сухого и смоченного термометров.

Существуют исследования, добавляющие учёт скорости ветра как множителя [31]. Комплексный показатель до сих пор находится в использовании у служб, занимающихся тушением пожаров, и ведется в течение всего теплого периода года ежедневно со дня схода снежного покрова до его выпадения. Однако важно отметить, что показатель не принимает в учёт количество осадков, которого может быть недостаточно в накопившемся итоге при одновременном выпадении осадков более 3 мм. Об этом также свидетельствует динамика изменения порогового значения с 5 мм, вызванная нестабильностью моделирования.

Для компенсации данной ситуации было предложено использовать показатели влажности напочвенного покрова ПВ-1 и подстилки ПВ-2 [32], расчет которых осуществляется по формулам 1.3 и 1.4:

$$\begin{cases} (\text{ПВ} - 1)_n = [(\text{ПВ} - 1)_{n-1} + t_{n-1}(t_{n-1} - r_{n-1})]K_n & \text{при } t > 0^\circ\text{C}, \\ (\text{ПВ} - 1)_n = [(\text{ПВ} - 1)_{n-1} + (t_{n-1} - r_{n-1})]K_n & \text{при } t \leq 0^\circ\text{C}, \end{cases} \quad (1.3)$$

$$\begin{cases} (\text{ПВ} - 2)_n = [(\text{ПВ} - 2)_{n-1}]K_n' + t_{n-1}(t_{n-1} - r_{n-1})K_n'' & \text{при } t > 0^\circ\text{C}, \\ (\text{ПВ} - 2)_n = [(\text{ПВ} - 2)_{n-1}]K_n' + (t_{n-1} - r_{n-1})K_n'' & \text{при } t \leq 0^\circ\text{C}, \end{cases} \quad (1.4)$$

где K_n, K_n', K_n'' – коэффициенты учёта осадков.

После расчета показателей пожарной опасности происходит деление на классы, отражающие различные вероятности возникновения пожаров и основанные на шкале Нестерова с определенными граничными значениями.

Полученное значение определяет класс пожарной опасности, к которой будет отнесен участок (таблица 1.1).

Таблица 1.1 – Зависимость классов пожарной опасности от комплексного показателя [33, с. 98]

Класс пожарной опасности	Величина комплексного показателя	Степень пожарной опасности
I	<300	Очень малая
II	301-1000	Малая
III	1001-4000	Средняя
IV	4001-12000	Высокая
V	>12000	Чрезвычайно высокая

I класс пожарной опасности присваивается территориям, на которых на данный момент времени природные условия и человеческая деятельность при обычных условиях не создаст серьезную угрозу быстрому распространению огня. V класс говорит о высокой вероятности возникновения как низовых, так и верховых пожаров и о высокой скорости их распространения.

Так, эти методики и шкала наиболее часто используются для определения пожарной опасности службами, занимающимися пожаротушением лесных массивов. Однако при этом распределение допускает наличие в I классе пожарной опасности до 5% случаев реального возникновения пожара от общего количества. В таблице 1.2 представлены средние значения числа пожаров по категориям природной пожарной опасности (КППО), произошедших в условиях определенного класса пожарной опасности за год для трех показателей в Свердловской области за последние 12 лет.

Можно заметить, что процентное соотношение не выполняется – доля пожаров на территориях с минимальной пожарной опасностью достаточно высока. Лишь для показателя ПВ-2 число пожаров I класса опасности достигает 6% среди остальных классов, но при этом доля пожаров V категории

чрезвычайно высока, что свидетельствует о возможном ложном срабатывании, которое ведет к созданию лишней нагрузки на работников и неправильному планированию.

Таблица 1.2 – Среднее число реальных пожаров с категориями

Показатель	КППО I	КППО II	КППО III	КППО IV	КППО V
Нестерова	102	117	167	64	90
ПВ-1	134	175	146	48	37
ПВ-2	34	55	124	73	254

Таким образом, эмпирические модели широко применяются благодаря своей относительной простоте и меньшим требованиям к вычислительным ресурсам по сравнению с теоретическими моделями. Они позволяют быстро получать результаты, что особенно важно при оперативном прогнозировании и принятии решений в условиях реальных пожаров. Однако для повышения их точности необходимо регулярно обновлять и дополнять базу данных новыми исследованиями и наблюдениями, а также существует необходимость в адаптации их к конкретным условиям различных регионов и к динамически меняющейся климатической ситуации в целом

1.3.2 Аналитические методы моделирования пожарной опасности

Теоретические модели основываются на принципах газовой динамики, математическом описании процессов горения, уравнении тепломассопереноса и других фундаментальных физических и химических законах. Обычно они представляются в виде системы дифференциальных уравнений в частных производных. Верификация подобных расчетов представляется непростой задачей, однако они могут описывать развитие лесного пожара с учетом общих и региональных факторов, а также текущего состояния окружающей среды, что позволяет решать широкий спектр вопросов в общем виде.

В отличие от классических расчётов, основанных на эмпирических приближениях, современные математические модели содержат в своей основе сложные алгоритмы для более точного и надежного прогнозирования. Они позволяют учитывать широкий спектр факторов, влияющих на возникновение и распространение пожаров, что делает их эффективным инструментом для специалистов по управлению рисками и чрезвычайным ситуациям.

В работе А.А. Кулешова [34] представлена следующая классификация математических моделей:

- Модели, прогнозирующие динамику распространения лесного пожара;
- Модели геометрической оценки лесного пожара;
- Модели предсказания характеристик течения, тепло- и массопереноса на кромке пожара и в самой зоне горения;
- Модели, прогнозирующие прочие характеристики.

Полуэмпирическая модель Ричарда Ротермела [35] является широко применяемым инструментом в области моделирования распространения низовых лесных пожаров. Она позволяет спрогнозировать скорость распространения фронта возгорания в направлении ветра в зависимости от характеристик воспламенившейся растительности, особенностей местности и скорости ветра. Эта модель обрела широкую программную реализацию [36, 37] и используется в лесных службах Северной Америки и Европы, несмотря на недостатки, связанные с отсутствием явного учета законов сохранения массы и импульса, что снижает точность и прогноза скорости.

Модель лесных пожаров, разработанная А. М. Гришиным [38], является одним из наиболее детально проработанных решений с физико-математической точки зрения. Её главная особенность – способность моделировать различные режимы горения на разных ярусах. Данная модель показывает значительную эффективность и высокую корреляцию с реальными последствиями пожара, однако вследствие использования большого числа параметров имеет высокую расчетную сложность. Ряд работ демонстрирует активное использование и дальнейшее развитие модели Гришина. Д.В.

Баровик уточнял геометрические параметры распространения верховых пожаров [39], а В.А. Перминов описывал процесс объединения множественных очагов горения [40].

Математическому моделированию возникновения лесных пожаров посвящено большое число современных работ, как отечественных [41-42], так и иностранных [43-44]. Несмотря на различные подходы и объекты исследования, в проектах подчеркивается широкая доступность новых источников данных, открывающих новые перспективы для развития и уточнения моделей.

Таким образом, анализ теоретических методов моделирования лесной пожарной опасности показывает их значительный потенциал для точного отражения процессов распространения огня в лесной местности. Эти модели позволяют предсказывать направление и скорость распространения пожаров с высокой степенью достоверности, что является важным для эффективного управления пожарами и защиты населенных пунктов.

Однако следует отметить, что вычислительные расходы на проведение расчетов в таких моделях могут быть значительными, особенно при моделировании в реальном времени или на больших территориях. Это ограничивает их практическое применение в оперативном прогнозировании и быстром реагировании на возможные пожары. Кроме того, подобные решения сосредоточены на распространении существующего очага возгорания и мало проработаны для прогнозирования возникновения новых точек.

1.3.3 Методы прогнозирования пожарной опасности на основе алгоритмов машинного и глубокого обучения

В предыдущих частях работы были рассмотрены математические модели развития пожаров, основанные на физических и эмпирических данных, которые позволяют оценить динамику распространения огня в различных условиях среды. Такие модели обеспечивают определённый

уровень понимания и предсказуемости поведения огня, однако они часто ограничены в способности учитывать сложные взаимодействия множества факторов и требуют долгих вычислений. По этой причине растет интерес к разработке и применению адаптивных и гибких методов прогнозирования, способных учитывать множество факторов и их взаимосвязи. Рассмотрим использование алгоритмов машинного и глубокого обучения в задачах прогнозирования пожарной опасности.

Существует несколько основных подходов к прогнозированию пожаров. Один из них включает анализ временных рядов, где модели используются для анализа исторических данных о погоде, влажности, ветровых условиях и других факторов. Такие алгоритмы способны учитывать зависимости и изменения условий во времени, что позволяет делать точные и своевременные прогнозы.

Для анализа временных рядов зачастую используется несколько известных архитектур нейронных сетей. Рекуррентные нейронные сети (RNN) — это вид искусственных нейронных сетей, специально разработанный для работы с последовательностями данных с возможностью продолжительного хранения информации о предыдущих состояниях. Технология LSTM (Long Short-Term Memory), относящаяся к этому виду нейросетей, специально предназначена для анализа временных рядов с учетом их долгосрочных зависимостей, что активно используется в ряде работ [45, 46, 47]. Другие научные труды [48] посвящены тестированию иных архитектур, также позволяющих проводить оценку временных рядов, в том числе управляемых рекуррентных блоков (Gated Recurrent Unit) и нейронной сети с задержкой по времени (Time Delay Neural Network).

Свёрточные нейронные сети (CNN), обычно ассоциирующиеся с анализом изображений, также могут быть адаптированы для решения задачи прогнозирования пожароопасности. Использование сверточных слоев позволяет выявлять повторяющиеся шаблоны в данных, что может быть полезно для прогнозирования пространственного поведения пожара [49].

Сочетание пространственных и временных данных позволяет моделям более точно предсказывать риск возникновения и распространения пожаров, поэтому модификация этих двух сетей, например, сверточная нейросеть долговременной памяти (ConvLSTM) также часто встречается в научных работах как метод для прогнозирования [50].

Помимо нейронных сетей, в прогнозировании пожаров эффективны ансамблевые методы машинного обучения, среди которых особое место занимают алгоритмы градиентного бустинга. В некоторых исследованиях сравнительный анализ показал, что модели градиентного бустинга показывают высокую эффективность без необходимости создания сложных алгоритмов и систем [51, 52].

Такие модели машинного обучения позволяют анализировать сложные и многомерные данные, объединяя множество слабых моделей, зачастую деревьев решений, в одну сильную модель. Принцип работы алгоритма будет описан далее – в подразделе 2.1.1. Этот подход к моделированию находит свое применение в практических приложениях. В качестве примера можно выделить проект с моделью `no_fire_with_ai` [53], позволяющей предсказывать пожароопасность на территории России при помощи бинарной классификации. Одним из главных преимуществ модели является проработанная географическая основа, учитывающая расположение населенных пунктов, объектов инфраструктуры и водоемов нашей страны. Далее более детально рассмотрим основу модели и ее применение в рамках данной задачи.

2 ВАЛИДАЦИЯ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ПОЖАРООПАСНОСТИ И ОПРЕДЕЛЕНИЕ КРИТЕРИЕВ, ПОВЫШАЮЩИХ ЭФФЕКТИВНОСТЬ ЕЕ РАБОТЫ

Эффективное управление рисками, связанными с пожарами, требует использования передовых моделей прогнозирования, способных предсказывать вероятность возникновения пожаров с высокой точностью, поэтому доработка имеющейся модели прогнозирования пожарной опасности с учетом других наработок может повысить ее точность и надежность, а также адаптировать модель к специфическим условиям различных регионов.

В данной главе будет проведен детальный анализ принципов работы выбранной модели. Будет рассмотрено, какие данные и алгоритмы лежат в её основе, как модель обрабатывает информацию и как строит прогнозы. Это позволит понять сильные и слабые стороны модели, а также её потенциал для адаптации и улучшения.

Затем необходимо провести валидацию модели, то есть оценить точность и надёжности ее предиктивной работы по предсказанию вероятности и локализации пожаров. После этого следует выявить направления для развития и улучшения, а также определить дополнительные факторы, которые не были учтены ранее – в том числе расширить набор данных, дополнить модель новыми параметрами. Проведенный ранее анализ факторов поможет разработать рекомендации для повышения точности и адаптивности модели, что в конечном счете улучшит её способность прогнозировать вероятность возникновения пожаров.

2.1 Анализ принципов работы модели

В данной работе рассматривается модель прогнозирования возникновения лесных пожаров, целью которой является определение вероятности возникновения пожара в каждой клетке территории страны, размером 0.2 на 0.2 градуса на горизонте прогнозирования от отчетной даты.

Основные этапы работы алгоритма включают в себя сохранение климатических данных, их предобработку и непосредственно процесс обучения и инференса модели. Рассмотрим подробнее каждую стадию.

2.1.1 Методы градиентного бустинга

Как описано в работе [54], задача градиентного бустинга сводится к построению оператора, который сможет как можно более точно предсказывать метку y из пространства Y для объекта x , принадлежащего многомерному пространству X . Обучающая выборка в таком случае состоит из объектов $\{x_i\}_{i=1}^N$ с метками $\{y_i\}_{i=1}^N$. Также имеются некоторые базовые алгоритмы H количеством M , которыми зачастую являются решающие деревья, при чем каждый элемент $h(x; a)$ определяется параметром a и входным вектором признаков x для каждого наблюдения. Для задачи классификации алгоритмы h передают степень уверенность принадлежности объекта классу.

Результат прогноза конечной модели представляет собой сумму слабых моделей, то есть, базовых алгоритмов, в соответствии с формулой 2.1:

$$F(x) = \sum_{m=1}^M h(x; a_m). \quad (2.1)$$

Для решения задачи подбора весов вводится функция потерь $L(y_i, F_m(x_i))_i$, показывающая, насколько предсказанный ответ отличается от правильного, как показано в формуле 2.2:

$$Q = \sum_{i=1}^N L(y_i, F_m(x_i)) \rightarrow \min. \quad (2.2)$$

Инициализация градиентного бустинга происходит константным значением $F_0(x_i) = \gamma, \gamma \in \mathbf{R}$. Тогда первый базовый алгоритм приближает решение к целевой переменной в соответствии с формулой 2.3:

$$h_0 = \underset{h \in H}{\operatorname{argmin}} \sum_{i=1}^N L(y_i, F_m(x_i)). \quad (2.3)$$

Следующие алгоритмы берут за основу для обучения поправку, определяющую разность между полученным предыдущим алгоритмом решением и правильной меткой. Эта поправка является производной от

функции потерь – антиградиентом определяется направление, в котором функционал ошибки уменьшается быстрее всего по формуле 2.4:

$$\nabla Q = - \left. \frac{\partial L(y_i, F_m)}{\partial F_m} \right|_{\partial F_m = h_k(x_i)} = h_k(x_i) - y_i \quad (2.4)$$

Далее рассмотрим применение алгоритма в задаче прогнозирования пожаров.

2.1.2 Сбор данных

Для задачи прогнозирования лесных пожаров с использованием алгоритма градиентного бустинга, необходимо адаптировать общие определения и концепции на конкретные параметры работы.

Признаки x из пространства X представляют собой различные климатические, географические и категориальные данные, которые могут влиять на возникновение лесных пожаров и будут использованы для обучения модели.

Климатические признаки извлекаются из GRIB-файлов (General Regularly distributed Information in Binary form), представляющих собой многомерные архивы метеорологических данных. Источником информации о температуре, скорости ветра, влажности, количестве осадков, испарении и тепловом излучении выступает набор ERA5-Land [55]. При этом, признаки имеют частоту изменений 0.1 на 0.1 градус в определенную дату и конкретной широте и долготы.

Географические признаки в работе во многом связаны с антропогенным воздействием человека: расстояние до ближайшего населенного пункта, его тип и численность, близость к дорогам, в том числе автомобильным и железнодорожным повышают пожароопасность, поскольку «вероятность возникновения пожара повышается в связи с неблагоприятной деятельности человека» [56, с. 50]. Города и поселения, как места постоянного скопления людей, представляют собой ключевой элемент антропогенной активности.

Помимо антропогенных факторов в географических признаках также учитывается наличие рек и других водоемов. Эти данные являются статическими для модели, их источник – OpenStreetMap [57].

Категориальные признаки в основном описывают временные зависимости и играют важную роль в прогнозировании пожаров, так как позволяют учитывать сезонные изменения, циклы активности человеческой деятельности и другие периодические факторы, которые могут влиять на вероятность возникновения и распространения пожаров в различные периоды времени. Это месяц, неделя с начала года, день недели, сезон. Мета-факторы, такие как индекс ячейки, дата не использовались в обучении, поскольку они описывают непосредственно объект, а не факт пожара. Однако в данных для обучения учитывается информация, описывающая наличие возгорания в самой ячейке или в определенном радиусе от нее за последнее время.

Каждый вид данных представляет собой отдельный признак x_i . Он является вектором с данными, описывающими параметры для каждой координатной ячейки, на которую разбивается территория страны. Метка y из пространства Y может принимать значения 0 или 1, где 0 обозначает отсутствие пожара, а 1 - его наличие. Далее рассмотрим подготовку этих данных.

2.1.3 Предобработка данных

В процессе разработки модели машинного обучения для прогнозирования пожарной опасности возникает необходимость обработки данных, которые имеют различную пространственную частоту. В данном случае исходные климатические данные для обучения имеют разрешение 0.1 на 0.1 градуса, тогда как прогнозы должны выполняться на сетке с разрешением 0.2 на 0.2 градуса. Для приведения данных к необходимому разрешению применяется техника *pooling features* – уплотнение карт признаков.

Пулинг помогает агрегировать информацию из более мелких ячеек в более крупные, сохраняя при этом существенную информацию для прогнозов. Кроме того, он обладает следующими преимуществами:

- Сокращение объема обрабатываемых данных: уменьшение разрешения данных снижает объем информации, что ускоряет обработку и анализ.

- Улучшение обобщающей способности модели: агрегация признаков помогает сгладить локальные вариации, делая модель менее чувствительной к шуму в данных.

Существует два типа пулинга: по максимальному значению (Max Pooling) в ячейке и по среднему (Average Pooling). В работе используются оба пулинга в качестве факторов по каждому признаку, поскольку максимальный позволяет выявлять повышенные значения показателей, которые могут способствовать возгоранию, а средний – сглаживает влияние отдельных аномальных измерений, делая данные более устойчивыми к выбросам и ошибкам в обработке данных.

Категориальные признаки также проходят предобработку, но уже средствами самой модели CatBoost, используя упорядоченное кодирование. Для этого метода доступны 4 типа: Borders (формула 2.5), Buckets (формула 2.6), BinarizedTargetMeanValue (2.7) и Counter (28).

$$category_i = \frac{CountInClass + prior}{TotalCount + 1}, \quad (2.5)$$

где $CountInClass$ – количество объектов с текущим значением категориального признака, для которых значение метки превышает i . Учитываются только те объекты, для которых уже рассчитано это значение,

$TotalCount$ – общее количество предшествовавших объектов, которые имеют значение признака, совпадающее с текущим,

$prior$ – константа.

$$category_i = \frac{CountInClass + prior}{TotalCount + 1}, \quad (2.6)$$

где $CountInClass$ – количество объектов с текущим значением категориального признака, для которых значение метки равно i . Учитываются только те объекты, для которых уже рассчитано это значение,

$TotalCount$ – общее количество предшествовавших объектов, которые имеют значение признака, совпадающее с текущим,

$prior$ – константа.

$$category_i = \frac{CountInClass + prior}{TotalCount + 1}, \quad (2.7)$$

где $CountInClass$ – сумма значений меток, деленная на максимальное целое число значений меток,

$TotalCount$ – общее количество объектов с аналогичным значением признака,

$prior$ – константа.

$$category_i = \frac{CurCount + prior}{MaxCount + 1}, \quad (2.8)$$

где $CurCount$ – общее количество объектов в наборе данных с текущим значением категориального признака,

$MaxCount$ – количество объектов с наиболее частым значением признака,

$prior$ – константа.

Таким образом, значение категориального признака перестает зависеть от метки и получает свой числовой целочисленный признак. Упорядоченное кодирование не создает лишних зависимостей и дополнительных столбцов, что значительно снижает размерность и увеличивает точность результата.

2.1.4 Обучение модели

После завершения этапов сбора и предобработки данных следующим важным этапом в разработке модели прогнозирования пожаров является ее обучение.

Модель прогнозирования пожаров представляет собой алгоритм, целью которого является предсказание вероятности возникновения пожара на заданную дату для каждой ячейки сетки, на которую разбита территория. Базовый алгоритм выдает прогноз на отдельные дни на неделю вперед, причем для каждого горизонта прогноза реализована отдельная бинарная классификация, сдвиг таргета которой позволяет учитывать информацию о предыдущих днях по аналогии с временными рядами. Такое решение позволяет добиться хорошей точности без необходимости создания сложных структур, как было описано в ряде работ [58, 59].

Для каждого дня недели t происходит инициализация собственной модели, определяемой с учетом предсказанной вероятности предыдущего дня по формуле 2.9.

$$F_M^t(x) = \sum_{m=1}^M b_m h(x^t y^{t-1}; a_m). \quad (2.9)$$

Затем происходит вычисление функции потерь и ее градиента, на основе которого обучается следующая модель в ансамбле в рамках одного дня t , как было описано в предыдущей главе. В итоге получается последовательность моделей, каждая из которых учитывает прогнозы предыдущих:

$$F(x) = [F^1(x), F^2(x'), F^3(x''), F^4(x'''), F^5(x^{iv}), F^6(x^v), F^7(x^{vi})]. \quad (2.10)$$

В качестве модели машинного обучения был выбран алгоритм CatBoost, относящийся к градиентному бустингу. Соответственно, в качестве базовых алгоритмов $h(x; a)$ используются решающие деревья.

Также алгоритм был предварительно доработан, целевым признаком стала вероятность возникновения пожара на недельный временной отрезок через заданный горизонт прогнозирования. При этом, при наличии пожара в нужные даты целевой признак будет равен 1 независимо от того, пожар только начался или продолжается уже некоторое время. Важно отметить, что модель не учитывает тушение пожара, поскольку эта задача зависит от множества сторонних факторов, не поддающихся предиктивной аналитике.

В качестве функции потерь использовалась логистическая функция, вычисляемая по формуле 2.11:

$$L(y, F_m(x)) = \log(1 + e^{-2yF_m(x)}). \quad (2.11)$$

Для получения качественного результата прогноза был проведен подбор гиперпараметров a для формулы 1.5 – характеристик модели, которые устанавливаются до начала процесса обучения и оказывают значительное влияние на её производительность. Процесс их подбора заключается в поиске таких значений, которые обеспечат наилучшую точность модели на валидационной выборке, а целью является нахождение оптимального баланса между переобучением и недообучением. Для градиентного бустинга выделяют следующие гиперпараметры:

- learning_rate – темп обучения,
- max_depth – глубина деревьев,
- n_estimators – количество базовых моделей,
- Другие гиперпараметры.

Набор гиперпараметров для лучшей модели приведен в таблице 2.1.

Таблица 2.1 – Результат подбора гиперпараметров

learning_rate	max_depth	n_estimators
0.01135	10	16933

Обучение проходило на сбалансированной выборке – процентное соотношение целевых переменных «пожар» и «без пожара» – 50:50. Метрикой при разработке был коэффициент Gini = 93.20. Этот показатель часто используется в задачах бинарной классификации для измерения способности модели различать два класса. Он не требовал выбора порога и хорошо подходил для сбалансированной выборки, которая была представлена для обучения. На несбалансированной выборке в реальных условиях такая метрика будет нестабильной из-за неустойчивости к редким событиям. По этой причине для интерпретации поведения модели необходимо рассчитать другие метрики.

2.2 Валидация работы модели на данных о реальных пожарах Российской Федерации

Важным этапом работы над моделью является валидация результатов работы на реальных данных. Этот процесс позволяет оценить точность и надёжность прогнозов, а также определить, насколько эффективно модель справляется с задачей в условиях, близких к реальным.

В данном разделе рассмотрим процесс валидации модели на данных о реальных пожарах, произошедших на территории Российской Федерации. Будут представлены методы и подходы, используемые для валидации, а также результаты, полученные в ходе проверки модели.

2.2.1 Изучение реальных данных

В качестве реальных данных будут использованы исторические данные о пожарах из датасетов MODIS, VIIRS [60] и МЧС [7].

Первый источник содержит развернутую информацию о термоточках в центрах пикселей радиусом 1 километр, которые были отмечены алгоритмами как ячейки, содержащие один или несколько очагов возгорания. Описание фиксируемых параметров для каждой термоточки представлены в таблице 2.2.

Таблица 2.2 – Поля параметров набора данных

№	Параметр	Описание
1	Широта	Координата X центра пикселя, причем не обязательно фактическое местоположение пожара, поскольку в пределах ячейки может содержаться более одного очага возгорания
2	Долгота	Координата Y центра пикселя, причем не обязательно фактическое местоположение пожара, поскольку в пределах ячейки может содержаться более одного очага возгорания

Продолжение таблицы 2.2

3	Яркостная температура	Яркостная температура огненного пикселя (в кельвинах)
4	Дата обнаружения	Дата обнаружения
5	Время обнаружения	Время обнаружения
6	Мощность излучения тепла	Отображает интегрированную в пикселях мощность излучения огня (в мегаваттах)
7	Время дня	День/ночь
8	Тип	0 - растительный пожар 1 - действующий вулкан 2 - другой статический наземный источник 3 - море

Так, термоточки типа «растительный пожар» на 07.06.2020-12.07.2020 отображены на рисунке 2.1 с помощью библиотеки matplotlib. За неделю зафиксировано 12544 точек природных возгораний.



Рисунок 2.1 – Данные о лесных пожарах из датасета MODIS

Во втором датасете, сформированном МЧС, из фиксируемых параметров, помимо широты, долготы и даты, выделяется тип пожара – контролируемый и неконтролируемый пал, лесной пожар, природный пожар, торфяной пожар. Термоточки этого набора данных представлены на рисунке 2.2. За неделю зафиксировано 4728 таких точек.



Рисунок 2.2 – Данные о лесных пожарах из датасета МЧС

Заметно, что данные не совпадают – это может быть связано как с более точным выбором природных пожаров по сравнению с обобщенными алгоритмами спутников, так и с занижением статистики пожаров в регионах. Сравним предсказания с обоими наборами данных.

2.2.2 Сравнение прогнозных значений с реальными данными

Изначально для валидации рассмотрим предиктивные результаты с горизонтом прогнозирования в 1 неделю. Распределение вероятностей возникновения пожаров на территории России на 07.06.2020-12.07.2020 представлено на рисунке 2.3.

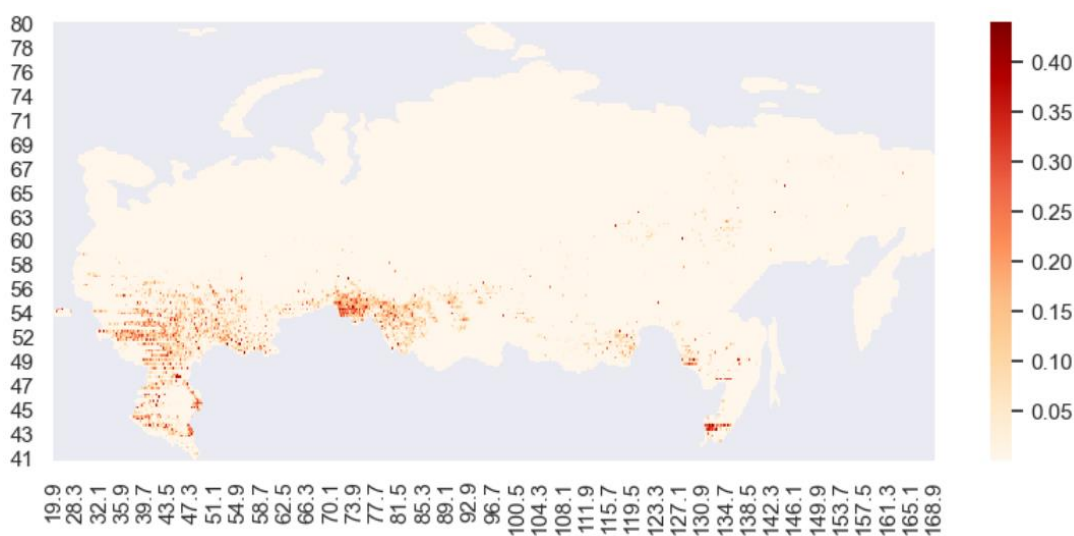


Рисунок 2.3 – Вероятности возникновения пожаров по результатам работы модели

По визуальной оценке, южная и западная части России относительно неплохо передаются моделью, однако необходим более конкретный анализ.

Поскольку спутниковые данные обладают большей точностью, чем шаг разбивки территории страны по сетке, необходимо корректно сопоставить данные различных точностей – целесообразно свести спутниковую информацию до порядка, используемого в модели прогнозирования. Для унификации данных применяется метод округления координат пожаров до ближайшего значения на сетке модели прогнозирования. Важно учитывать, что в таком случае возможна ситуация, при которой в одной ячейке сетки может оказаться несколько горячих точек.

Действительно, при объединении массивов прогнозных значений с реальными пожарами появились ячейки-дубликаты, свидетельствующие о наличии нескольких термоточек в одной ячейке. Группировка значений показала, что их число варьируется от 0 до 159 на клетку для датасета MODIS и от 0 до 38 для датасета МЧС. Расчет корреляции путем вычисления коэффициента Пирсона между столбцами DataFrame в библиотеке Pandas между количеством термоточек на одной ячейке и вероятностью возникновения пожаров не показал хорошего результата:

$corr_{mes} = 0.0003$ – для датасета МЧС,

$corr_{nasa} = 0.0019$ – для датасета MODIS.

Расчет зависимости между другими характеристиками термоточек приведен в таблице 2.3. Заметно, что у числа термоточек в ячейке имеется высокая корреляция с мощностью излучения, которая является тепловым потоком, характеризующим интенсивность теплоотдачи во время горения. То есть, суммарное тепловое излучение указывает на более частые или интенсивные пожары в данной области. Вероятность возникновения пожара же не имеет сильно выраженную зависимость ни от одной из характеристик.

Таблица 2.3 – корреляция вероятности возникновения пожара и параметров, фиксируемых у термоточек

	Вероятность пожара	Яркостная температура (21 канал)	Яркостная температура (31 канал)	Мощность излучения	Число пожаров в ячейке
Вероятность пожара	1.000000	0.042547	0.043676	0.007651	0.001871
Яркостная температура (21 канал)	0.042547	1.000000	0.997695	0.471869	0.426066
Яркостная температура (31 канал)	0.043676	0.997695	1.000000	0.421425	0.383108
Мощность излучения	0.007651	0.471869	0.421425	1.000000	0.775216
Число повторов	0.001871	0.426066	0.383108	0.775216	1.000000

2.2.3 Поиск доверительного порога

Поскольку модель определяет значение вероятности в интервале от 0 до 1, где 0 означает отсутствие риска пожара, а 1 – максимальный риск, имеет смысл найти доверительный порог, при значениях выше которого вводится режим повышенной готовности для реагирования. Правильный выбор уровня имеет критическое значение, так как слишком низкий порог может привести к ложным срабатываниям, в то время как слишком высокий вызовет пропуск реальных угроз. Это значение должно минимизировать количество ложноположительных и ложноотрицательных прогнозов, обеспечивая при этом высокую надежность и эффективность работы экстренных служб. Для этого используются различные метрики.

В качестве метрик для оценки моделей обычно используются следующие: Precision, Recall, Accuracy.

Precision (или точность) – метрика, которая показывает долю верно определенных прогнозов среди всех положительных прогнозов, сделанных моделью. Рассчитывается по формуле 2.12:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (2.12)$$

где TP (True Positives) – количество истинно положительных прогнозов,

FP (False Positives) – количество ложноположительных прогнозов, то есть неверно определенных моделью.

Высокая точность означает, что модель редко выдает ложные прогнозы, что важно для эффективного использования ресурсов пожарных служб.

Recall (или полнота) – метрика, которая показывает долю верно определенных положительных прогнозов среди всех реальных случаев пожара в соответствии с формулой 2.13:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (2.13)$$

где FN (False Negatives) – количество ложноотрицательных прогнозов, то есть прогнозов, в которых модель не смогла определить пожар.

Высокая полнота отражает факт того, что модель обнаруживает большинство случаев пожара, что важно для предотвращения как можно большего числа пожаров.

Accuracy – метрика, которая показывает долю правильных прогнозов, как положительных, так и отрицательных, среди всех прогнозов, по формуле 2.14.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (2.14)$$

где TN (True Negatives) – количество истинно отрицательных прогнозов, то есть прогнозов, где модель правильно определила отсутствие пожара.

Accuracy показывает общую эффективность модели, но может быть менее информативной в случае несбалансированных данных, например, когда таргетов наличия пожара значительно меньше, чем таргетов без пожара. В таких условиях оценка может быть завышенной из-за большого числа истинно отрицательных прогнозов даже при малом количестве истинно

положительных результатов. По этой причине использовать эту метрику в данном контексте нецелесообразно.

Precision и Recall одинаково важны в задаче прогнозирования пожара, поскольку необходимо избегать как ложных реагирований для минимизации ненужного развертывания ресурсов, так и недостатка внимания для предотвращения недооценки угрозы. В таком случае эффективно использовать метрику F1-score – среднее гармоническое между точностью и полнотой, которое применяется для учета как ложноположительных, так и ложноотрицательных прогнозов в соответствии с формулой 2.15.

$$F1 = 2 * \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.15)$$

Так, для определения доверительной границы необходимо составить матрицу ошибок для различных порогов и определять прогнозные значения положительными, если они выше этой границы, и отрицательными в ином случае. Затем необходимо сравнить с фактическими пожарами прогнозы, выявить ложные и истинные срабатывания и рассчитать метрики. Результат представлен на рисунке 2.4.

Максимальное значение метрики F1 получается при вероятности пожара 0.15, что является плохо трактуемым результатом.

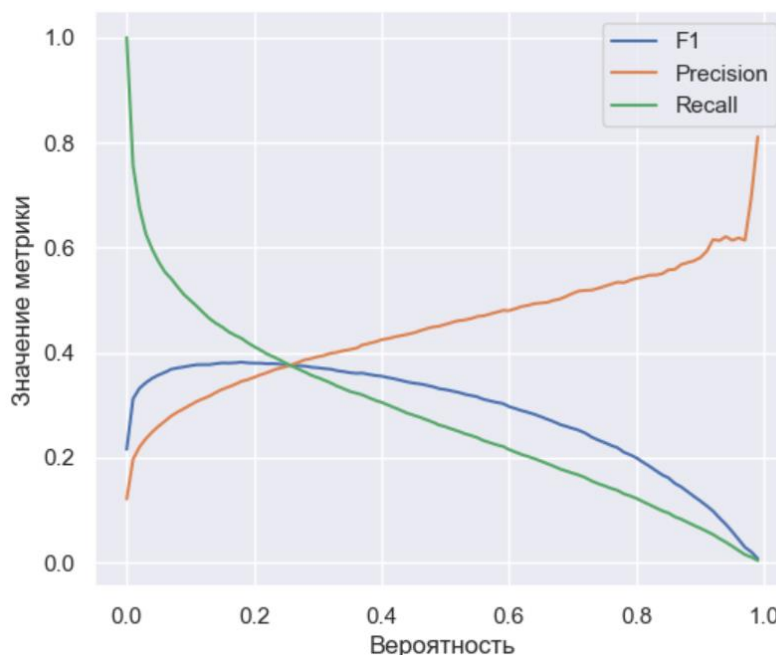


Рисунок 2.4 – Метрики, оценивающие работу модели

В задачах прогнозирования пожаров важно не только оценивать качество модели по отдельным метрикам, таким как точность и полнота, но и иметь возможность визуально и количественно анализировать ее поведение при различных пороговых значениях. В этом контексте кривая ROC (Receiver Operating Characteristic) и площадь под кривой ROC (AUC - Area Under Curve) играют ключевую роль.

ROC-кривая – графическое представление, которое позволяет оценить соотношение между метрикой Recall, и долей ложноположительных результатов при различных значениях порогов, вводимых для отнесения объекта к тому или иному классу по итогам работы модели. AUC – числовая оценка, представляющая площадь под ROC-кривой. Чем ближе ее значение к 1, тем лучше модель находит отличия между положительными и отрицательными классами.

Получившаяся кривая изображена на рисунке 2.5.

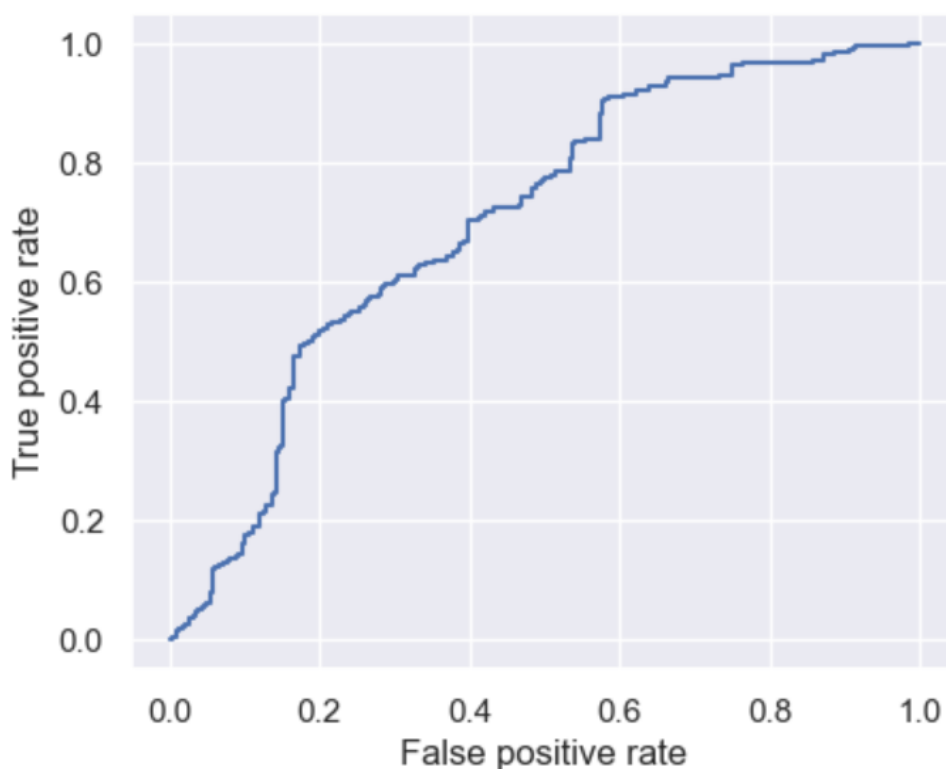


Рисунок 2.5 – ROC-кривая

ROC-AUC = 0.7037, что является удовлетворительным результатом, поскольку допустимый уровень ошибки прогнозирования развития пожаров составляет 30–35 процентов [42].

Результаты валидации позволяют сделать вывод о применимости модели для реальных задач прогнозирования пожаров, однако для улучшения ее эффективности и повышения надежности прогнозов в различных условиях эксплуатации рекомендуется провести дополнительные исследования, в том числе для ее адаптации к специфическим условиям различных регионов. Как минимум, необходимо представить более развернутую информацию о характеристиках растительного покрова и подстилающей поверхности, состояния и статуса эксплуатации леса. В следующей главе рассмотрим создание алгоритма машинного обучения, учитывающего данные характеристики.

3 УЛУЧШЕНИЕ РЕШЕНИЯ И ТЕСТИРОВАНИЕ ГИПОТЕЗ НА ПРИМЕРЕ ОДНОГО ИЗ РЕГИОНОВ РОССИИ

В предыдущей главе была детально рассмотрена имеющаяся модель машинного обучения для прогнозирования пожароопасной ситуации. Некоторые метрики показали хорошие результаты, но, тем не менее, необходимы значительные доработки для повышения точности и адаптивности к специфическим условиям регионов.

Цель данной главы заключается в разработке новой модели машинного обучения, учитывающей недостатки описанного выше решения. Для более точной интерпретации данных было принято решение сузить рассматриваемую область до одного региона России. В качестве территории для исследований будет использована Свердловская область, что позволит получить более детализированные и локальные выводы. Выбор региона обусловлен доступностью данных и их репрезентативностью для исследуемой задачи.

Свердловская область – субъект Российской Федерации, расположенный в центральной части Урала. Лесистость области достигает 66,5%, при чем растительность в основном хвойная и лиственная, что представляет потенциальный риск возгорания. Регион имеет высокую плотность населения и развитую инфраструктуру.

Рассмотрим, как базовая модель прогнозирования пожарной опасности работает на территории Свердловской области за 2019 и 2023 года по ежемесячным данным (таблица 3.1). Для этого будем использовать метрику ROC-AUC, поскольку она не требует порогового значения, расчет которого нецелесообразен, так как в масштабах страны у региона стабильно низкий уровень пожароопасности: за 2019 год максимальное значение, предсказанное базовой моделью – 0.2334, а за 2023 – 0.38465. Также не будем рассматривать месяцы, когда пожары не происходили, поскольку расчет ROC-AUC в них невозможен.

Таблица 3.1 – Значение метрики ROC-AUC для базовой модели прогнозирования пожаров

	Апрель	Май	Июнь	Июль	Август	Сентябрь
Год	2019					
Число пожаров	19	167	24	47	3	1
ROC-AUC	0.65	0.69	0.46	0.79	0.73	0.91
Год	2023					
Число пожаров	108	206	124	121	50	61
ROC-AUC	0.73	0.66	0.72	0.64	0.65	0.63

Можно заметить, что точность модели нестабильна и значительно меняется от месяца к месяцу. При очень маленьком числе пожаров корректность работы модели также сложно оценивать. Для создания новой модели прогнозирования пожарной опасности необходимо учитывать следующие аспекты:

- Интеграция дополнительных факторов, учитывающих геоморфологические характеристики региона;
- Способность адаптироваться к изменениям в данных, в том числе к колебаниям количества пожаров в разные периоды времени.

Для решения последней задачи опробуем подход к определению зон пожарной опасности путем интерполяции данных о точках реальных пожаров за временные интервалы, предшествующие дате прогноза. Это позволит модели прогнозировать изменения на основе динамики числа пожаров и площади пожароопасности прошлых недель.

3.1 Построение карт пожарной опасности

Одним из ключевых аспектов управления пожарной опасностью является способность быстро и точно идентифицировать потенциально опасные зоны возникновения и распространения возгорания.

Как упоминалось в первой главе, характеристика классов пожарной опасности дает подробное представление об элементе возгорания и активно используется в службах мониторинга. Кроме того, географическое расположение этих классов может способствовать пространственному анализу и учету трендов развития пожароопасной обстановки региона. Так, на рисунке 3.1 представлено пространственное размещение пожаров, разбитых на классы по методике Нестерова, описанной в подразделе 1.3.1, за 2020 год в Свердловской области.

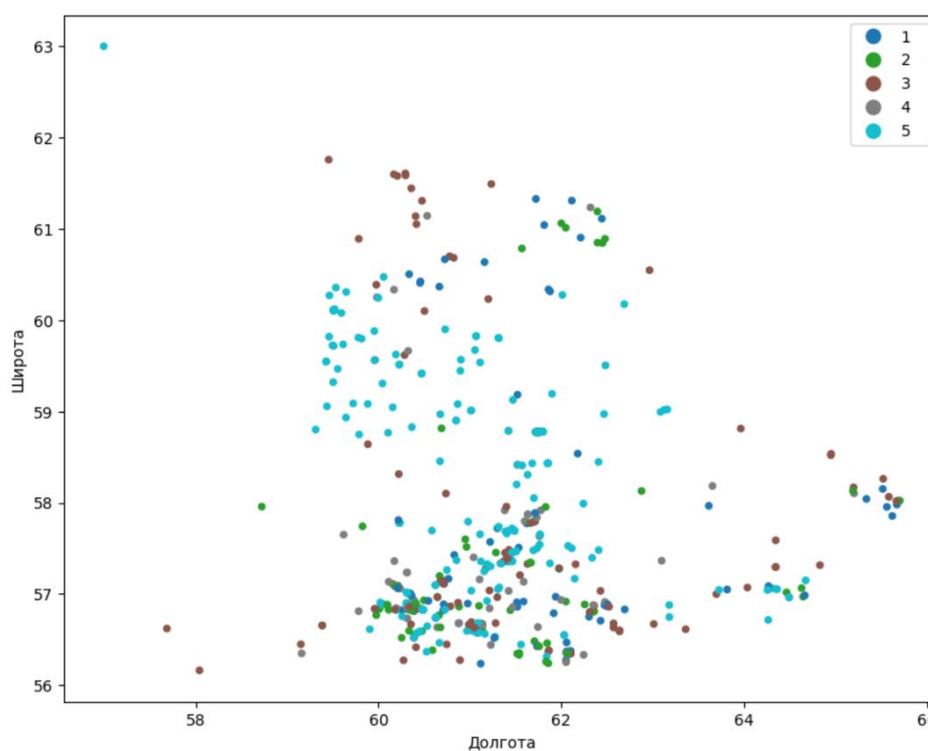


Рисунок 3.1 – Классы пожарной опасности возгораний за 2020 год

С целью создания более целостных и применимых на практике решений необходимо не только определять класс пожара, но и эффективно визуализировать и структурировать эти данные. Для этого необходимо перейти от индивидуальной классификации пожаров к объединению этих данных в более крупные структуры, характеризующиеся одинаковым уровнем пожарной опасности. Интерполяция позволит определить зоны пожарной

опасности на всей территории, основываясь на имеющихся данных о классификации произошедших пожаров.

Для анализа были применены два метода интерполяции: интерполяция радиальной базисной функции и триангуляция с линейной интерполяцией, поскольку они хорошо справляются с нерегулярными данными [61, 62].

Радиально-базисные функции (RBF) используются для интерполяции данных в многомерных пространствах путем создания сглаженных поверхностей. Интерполяционная функция $f(x, y)$ определяется по формуле 3.1:

$$f(x, y) = \sum_{i=1}^N w_i \phi(\| (x, y) - (x_i, y_i) \|), \quad (3.1)$$

где w_i – веса, определенные на основе значений в известных точках x_i, y_i ,

ϕ – радиально-базисная функция, которая зависит от расстояния между точкой (x, y) и опорной точкой (x_i, y_i) ,

N – количество известных опорных точек.

Под опорными понимаются точки, значения в которых известны – это точки пожаров с классами пожарной опасности. Функция *Rbf* из пакета *scipy* создает интерполяционную функцию на основе радиально-базисных функций, в данном случае – линейную. На границах устанавливаются значения первого класса пожарной опасности для стабилизации решения и предотвращения резких изменений на краях. Первый класс пожарной опасности не был учтен в интерполяции, поскольку при его близком расположении к другому классу территория принимает значение большего класса, а на территории вне высоких классов и так принимается значение, равное 1. Результаты для четырех месяцев 2020 года изображены на рисунке 3.2.

В целом, интерполяция с использованием радиально-базисных функций была выполнена корректно, учитывая следующие наблюдения:

- Графики показывают плавные и непрерывные изменения классов пожарной опасности.
- Интерполяция охватывает все области и корректно отображает значения классов.

- Установленные значения на краях сетки предотвращают резкие скачки и артефакты на границах графика.

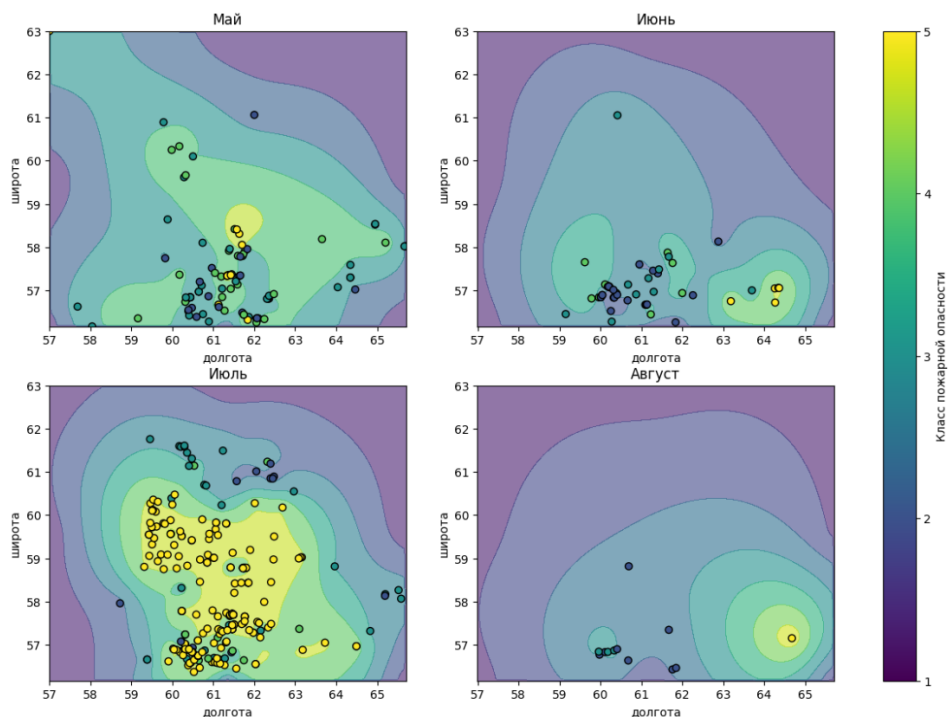


Рисунок 3.2 – Интерполяция классов пожарной опасности с помощью радиально-базисных функций для четырех месяцев 2020 года

Линейная интерполяция на основе триангуляции Делоне делит плоскость на множество треугольников таким образом, чтобы не было узлов внутри описанных окружностей (рисунок 3.3) [63]. Внутри каждого такого треугольника выполняется линейная интерполяция.

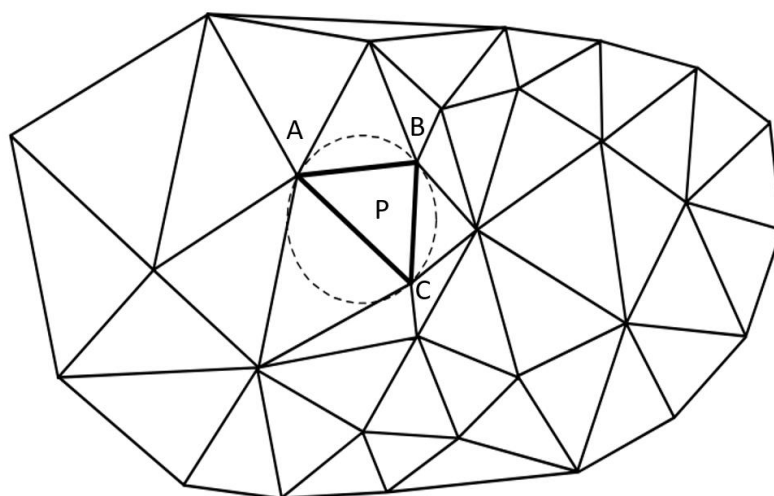


Рисунок 3.3 – Триангуляция Делоне

Для треугольника с вершинами A , B , C со значениями функции в вершинах $f(A)$, $f(B)$, $f(C)$ значение в точке P будет вычислено по формуле 3.2:

$$f(P) = \lambda_A f(A) + \lambda_B f(B) + \lambda_C f(C), \quad (3.2)$$

где λ_A , λ_B , λ_C – это барицентрические координаты точки P , рассчитываемые по формулам (3.3):

$$\lambda_A = \frac{S_{PBC}}{S_{ABC}}, \lambda_B = \frac{S_{PAC}}{S_{ABC}}, \lambda_C = \frac{S_{PAB}}{S_{ABC}}. \quad (3.3)$$

Функция *LinearNDInterpolator* из пакета *scipy* создает интерполяционную функцию на основе триангуляции. Значения в каждой точке сетки вычисляются с использованием линейной комбинации значений в вершинах треугольника, в который попадает эта точка. В районах с малой плотностью данных, особенно на краях области интерполяции, триангуляция может содержать большие треугольники, что приводит к резким изменениям интерполируемых значений, поэтому для стабилизации интерполяции и предотвращения артефактов граничные значения в углах карты были заданы первым классом пожарной опасности.

Результаты интерполяции для четырех месяцев 2020 года изображены на рисунке 3.4.

С неравномерной плотностью данных триангуляция создает резкие изменения интерполируемых значений, а также она плохо справляется в случаях разреженных значений, что можно заметить на рисунке 3.5. RBF-интерполяция обеспечивает гладкую интерполяцию, которую можно также настраивать изменением самих функций. Триангуляция Делоне же является кусочно-линейной функцией. Кроме того, для больших территорий с высокой размерностью, а также в случаях с большим числом точек пожаров разбишка карты на треугольники может потребовать значительных вычислительных ресурсов.

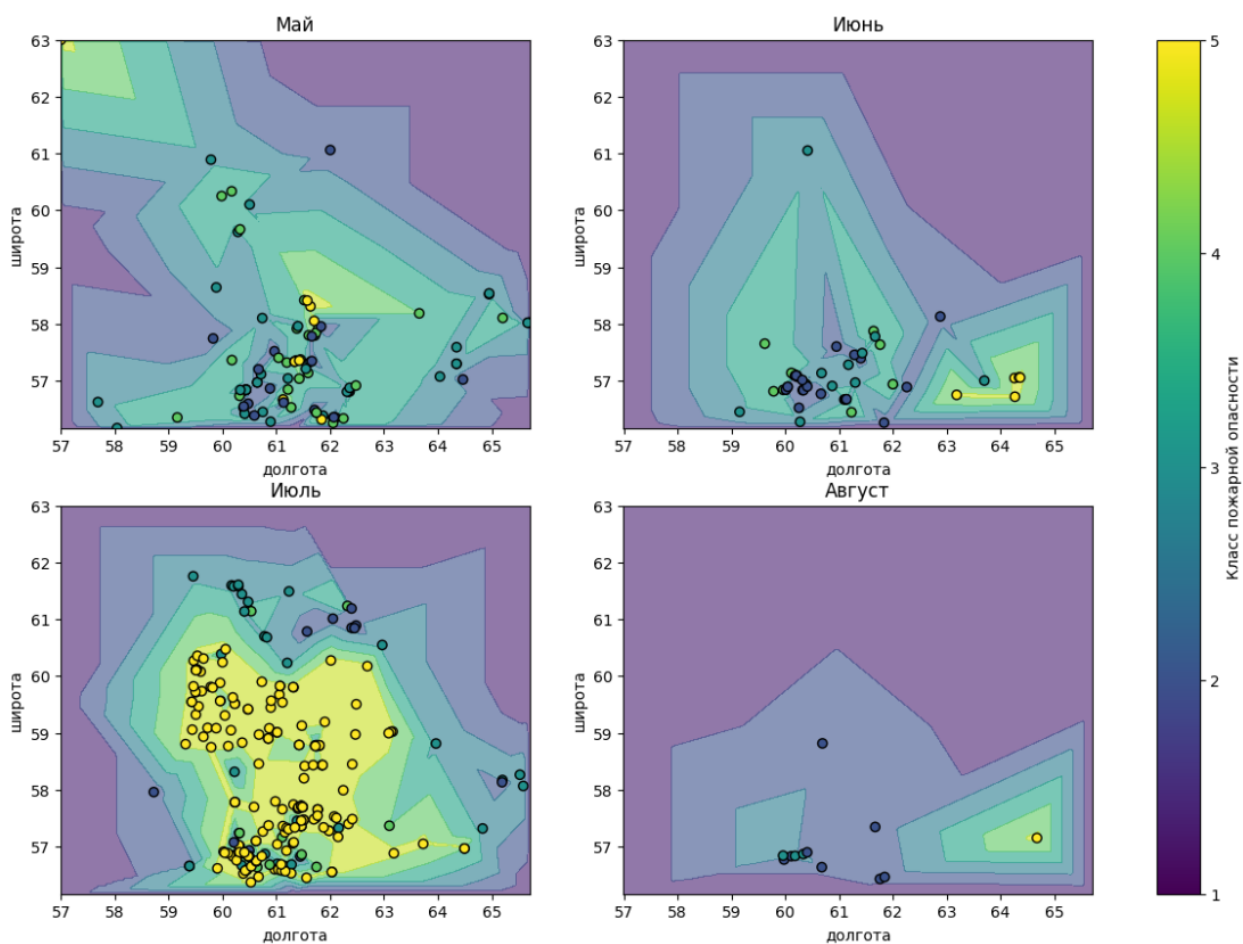


Рисунок 3.4 – Линейная интерполяция на основе триангуляции Делоне классов пожарной опасности для четырех месяцев 2020 года

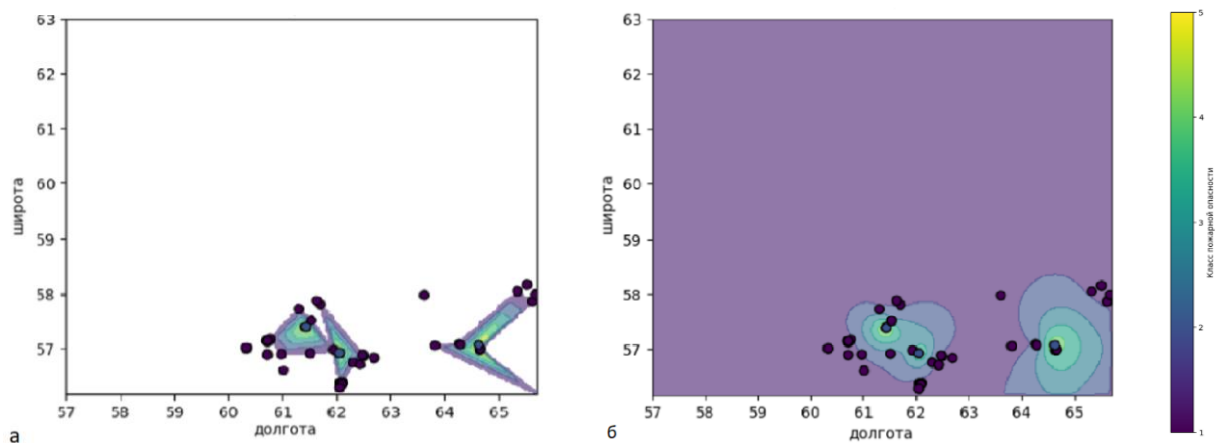


Рисунок 3.5 – Результаты интерполяции пожаров на апрель 2020 года
 (а) – Интерполяция с радиально-базисными функциями,
 (б) – Линейная интерполяция на основе триангуляции Делоне

Было принято решение дообучить исходную модель прогнозирования пожарной опасности на интерполированных радиально-базисными функциями картах классов пожаров предыдущих временных интервалов. Предполагается, что в динамике прогнозирования на малые горизонты планирования это может показать эффективность.

3.2 Разработка новой модели

В задаче подготовки данных для обучения модели важно отличать данные, которые можно использовать для составления прогноза. Так, важны параметры без прямого использования информации, характеризующей сами пожары, или данных, которые являются недоступными на момент составления прогнозов, например, классы пожарной опасности на неделю прогноза. Таким образом, в качестве параметров для каждой ячейки сетки, на которую разбивается территория, были выбраны:

- Предсказание пожароопасности базовой модели,
- Класс пожарной опасности на 1, 2, 3 недели до прогнозной недели,
- Уточненные геоморфологические факторы, которые были недостаточно подробно учтены в базовой модели. К ним относятся тип и состояние растительности и покрова почвы, категория земель статус эксплуатации территории.

Различные климатические факторы подробно заложены как в базовую модель, так и в ранжировку по классам, поэтому было принято решение не использовать их повторно.

В качестве модели обучения был выбран *CatBoostClassifier* со следующими гиперпараметрами, полученными в результате подбора (таблица 3.2).

Таблица 3.2 – Значения гиперпараметров

learning_rate	max_depth	n_estimators	l2_leaf_reg
0.15322	7	500	1.73739

Для корректного сравнения с базовой моделью было произведено обучение новой модели на пожарах 2020 года с последующим тестированием на 2019 году в соответствии со схемой на рисунке 3.6, причем в качестве выходного значения использовалась вероятность возникновения пожара с помощью функции `predict_proba`.



Рисунок 3.6 – Структура разбиения данных на обучающую, валидационную и тестовую выборки

Число пожаров за каждую неделю 2019 года можно оценить на рисунке 3.7. Представлены недели, в которых происходил как минимум один лесной пожар.

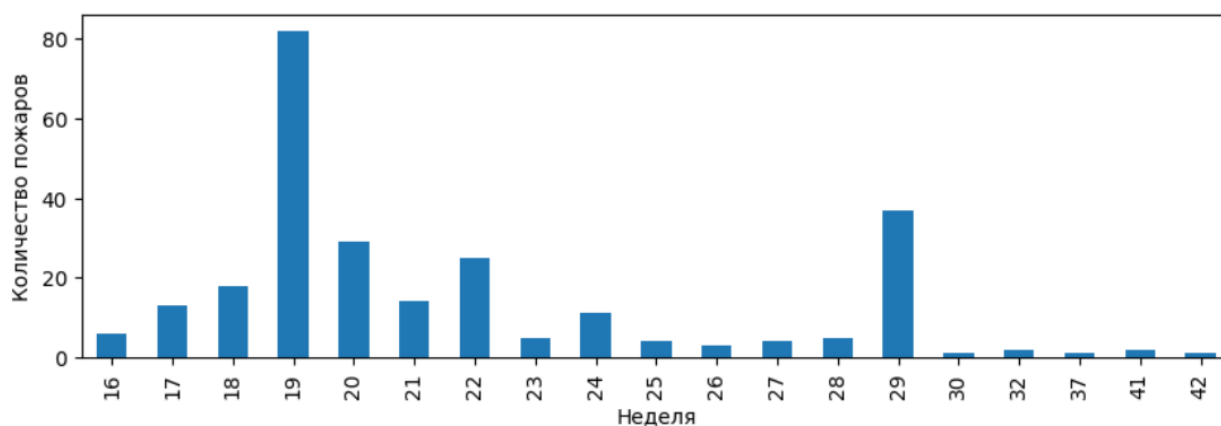


Рисунок 3.7 – Количество пожаров по неделям в 2019 году

Результаты сравнения метрики ROC-AUC новой модели с базовой для недель с высокой пожарной опасностью представлены в таблице 3.3.

Таблица 3.3 – Сравнение результатов базовой и дообученной модели

№ недели	19	20	21	22	23	24
Базовая модель, ROC-AUC	0.691	0.768	0.728	0.581	0.463	0.574
Новая модель, ROC-AUC	0.823	0.852	0.813	0.851	0.801	0.711
№ недели	25	26	27	28	29	30
Базовая модель, ROC-AUC	0.932	0.34	0.466	0.793	0.774	0.67
Новая модель, ROC-AUC	0.990	0.963	0.846	0.963	0.764	0.9

Анализ показывает, что новая модель имеет более высокие значения ROC-AUC на всем периоде. Для более точного анализа рассчитаем средние значения ROC-AUC для каждой модели и вычислим прирост ROC-AUC. Так, средние значения метрик ROC-AUC равны 0.648 для базовой модели и 0.856 для новой. ROC-кривые для базовой и новой модели представлены на рисунке 3.8.

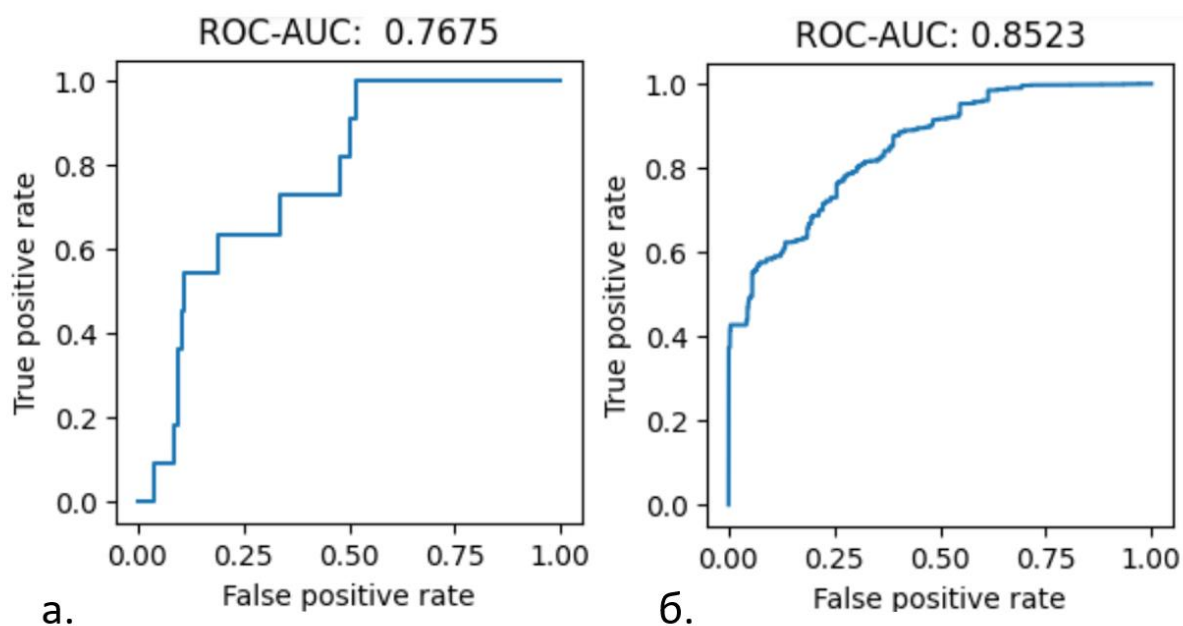


Рисунок 3.8 – ROC-кривые на 20-ю неделю 2019 г.

(а) – Базовая модель,

(б) – Новая модель

Кривая ROC базовой модели имеет ступенчатую форму, что может свидетельствовать о менее плавной работе на различных порогах. Более того,

при одинаковых значениях false positive rate чувствительность старой модели ниже, что указывает на её менее уверенную работу по сравнению с новой моделью. Новая модель дает меньшее количество ложноположительных и ложноотрицательных предсказаний. Таким образом, получившаяся модель обладает стабильно более высокой надежностью по сравнению с базовой в сезон высокой пожарной опасности.

Рассмотрим также метрики прогнозов на месяцы с низкой пожарной опасностью – для их обучения карты интерполяции имеют малое число точек или не имеют их вовсе – в последнем случае класс пожарной опасности по всей территории задается равным 1. Результаты прогнозирования представлены в таблице 3.4.

Таблица 3.4 – Сравнение результатов базовой и новой модели

№ недели	17	18	32
Базовая модель, ROC-AUC	0.716	0.764	0.732
Новой модель, ROC-AUC	0.904	0.773	0.871

В таких условиях качество работы модели менее стабильное. Для вариантов, когда карты пожарной опасности всех предшествующих периодов не имеют ни одной точки возгорания и полностью относятся к 1 классу пожарной опасности, новая модель не прогнозирует их появление. Таким образом, модель не может быть использована для предсказания начала пожароопасного периода.

Итак, в данной главе на примере Свердловской области было проведено тестирование базовой версии модели, показавшее необходимость ее улучшения. Для этого были созданы еженедельно обновляемые карты пожарной опасности региона на основе интерполяции классов пожарной опасности реальных очагов возгораний. В процессе работы был выбран оптимальный алгоритм интерполяции, который позволил наиболее точно отобразить распределение пожаров и потенциально опасных зон. Затем, учитывая сильные и слабые стороны базовой модели, была реализована новая

модель с учетом ряда дополнительных параметров. В их число вошли данные, полученные с динамических карт, а также геоморфологические факторы, описывающие характеристики растительности и рельефа. Такой подход позволил учитывать местные особенности региона, влияющие на возникновение и распространение пожаров. Новая модель показала более высокие значения ROC-AUC по сравнению с базовым алгоритмом, что свидетельствует о повышении точности прогнозирования вероятности возникновения лесных пожаров, особенно для временных периодов с высокой пожарной активностью.

Заключение

Основные результаты работы заключаются в следующем:

- В результате анализа литературы выделены основные факторы, влияющие на возникновение и развитие лесного пожара. Проанализированы имеющиеся подходы к прогнозированию пожароопасности, в том числе эмпирические, теоретические и основанные на алгоритмах машинного обучения.

- Описан алгоритм градиентного бустинга для задачи прогнозирования лесных пожаров. Изучена и проанализирована модель *no_fire_with_ai* на применимость к прогнозированию пожаров на территории России и отдельно в Свердловской области: рассчитаны метрики и произведен выбор доверительного порога. Результаты указали на необходимость модификации модели для улучшения точности прогнозов.

- На основе анализа недостатков существующей модели была разработана новая модель. Наряду с факторами, которые лежат в основе базовой модели, новый алгоритм учитывает характеристики местности и растительности рассматриваемого региона, а также пространственное распределение опасных зон возникновения и распространения возгораний. Для этого был опробован подход к определению зон пожарной опасности региона за временные интервалы, предшествующие дате прогноза, путем интерполяции данных о точках реальных пожаров с целью создания карт пожарной опасности. Таким образом, новая модель учитывает не только факты возгораний, но и рассматривает потенциально пожароопасные зоны в целом. Подход показал высокую эффективность, повысив точность прогнозирования пожаров новой модели по сравнению с базовой.

Список использованных источников

1. Influences of wildfire on the forest ecosystem and climate change: A comprehensive study / G. Kandasamy, K. Sabariswaran, N. Mathiyazhagan // *Environmental Research*. – 2024. – Vol. 240. – Part 2.
2. Climate Change, Forest Fires, and Health in California / R. Cisneros et al. // *Climate Change and Air Pollution*. Springer International Publishing. – 2017. – P. 99–130.
3. Хроника ООН: по мере роста числа природных пожаров все более актуальными становятся комплексные стратегии в области лесного хозяйства, климата и устойчивого развития // Организация Объединенных Наций : официальный сайт. – 2023. – URL: <https://www.un.org/ru/208325> (дата обращения: 20.05.2024).
4. Лапин, Д.Н., Аксенов, С.Г. К вопросу о лесных пожарах в российской федерации / Д.Н. Лапин, С.Г. Аксенов // *E-Scio*. – 2023. – №1 (76). – С. 15-20.
5. В 2023 году площадь лесных пожаров в России почти в 2 раза ниже среднепятилетних показателей. // Социальный фонд России : официальный сайт. – 2023. – URL: https://sfr.gov.ru/press_center/z_news~2023/11/17/257123 (дата обращения: 20.05.2024).
6. Roser, M., Ortiz-Ospina, E. Global Rise of Education // *Our World in Data*. – 2019. – URL: <https://ourworldindata.org/wildfires#all-charts> (accessed: 21.05.2024).
7. Пожары в России: данные о местах и типах природных пожаров за 2012–2021 гг. // МЧС; обработка: Новиков В.А., Тихонов С.В., Инфраструктура научно-исследовательских данных. АНО «ЦПУР», 2022. Доступ: Лицензия ССВУ-SA. – 2022. – URL: https://data.rcsi.science/data-catalog/datasets/202/#dataset-custom_tab_63 (дата обращения: 21.05.2024).
8. Курбатский, Н.П. Терминология лесной пирологии / Н.П. Курбатский // *Вопросы лесной пирологии*. – 1972. – С. 171-231.

9. Причины лесных пожаров // Уральская база авиационной охраны лесов : официальный сайт. – URL: <https://avialesural.ru/info/> (дата обращения: 21.05.2024).

10. Коровин, Г. Н. Влияние климатических изменений на лесные пожары в России / Г. Н. Коровин, Н. В. Зукерт // Климатические изменения : взгляд из России / под ред. В. И. Данилова Данильяна. – М. : ТЕИС. – 2003. – С. 69-98.

11. Кутлуахметов, Р.Р., Аксенов, С.Г. Анализ влияния изменения климата на природные пожары (на примере Российской Федерации и ряда зарубежных стран) / Р.Р. Кутлуахметов, С.Г. Аксенов// Экономика строительства. – 2023. – №9. – С. 92-94.

12. Петров, М. И. Влияние климатических условий на лесные пожары в Центральной Якутии / М. И. Петров, А. Н. Федоров // Природные ресурсы Арктики и Субарктики. – 2023. – Т. 28, № 2. – С. 248-260. – DOI 10.31242/2618-9712-2023-28-2-248-260.

13. Циркуляционные факторы возникновения лесных пожаров на территории Сибири и Дальнего Востока летом 2019 и 2021 гг. / И.В. Латышева, С.Ж. Вологжина, К.А. Лощенко // Известия Иркутского государственного университета. Серия: Науки о Земле. – 2021. – Т. 38. – С. 54-70.

14. Changing wildfire, changing forests: the effects of climate change on fire regimes and vegetation in the Pacific Northwest, USA / J.E. Halofsky, D.L. Peterson, B.J. Harvey // Fire ecology. – 2020. – №16. – Vol. 4.

15. Wildfire response to changing daily temperature extremes in California's Sierra Nevada / A.A. Gutierrez et al. // Science Advances. – 2021. – Vol. 7. – №47.

16. Global and Regional Trends and Drivers of Fire Under Climate Change / Jones M.W. et al. // Reviews of Geophysics. – 2022. – Vol. 60. – № 3.

17. Кузнецова, Е. Ю. Влияние экологических последствий лесных пожаров на изменение климата / Е. Ю. Кузнецова // Пожарная безопасность: современные вызовы. Проблемы и пути решения : Материалы Всероссийской научно-практической конференции, Санкт-Петербург, 26 апреля 2022 года. –

Санкт-Петербург: Санкт-Петербургский университет Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий имени Героя Российской Федерации генерала армии Е.Н. Зиничева, 2022. – С. 188-190.

18. Бурлаков, П.С., Хмара, К.А. Лесные пожары от гроз как геоэкологический фактор устойчивости светлохвойных лесов бассейна реки Сояна на Беломорско-Кулойском плато / П.С. Бурлаков, К.А. Хмара // Труды КарНЦ РАН. – 2011. – №1.

19. White, R. H. Effect of Lignin Content and Extractives on the Higher Heating Value of Wood / R. H. White // Wood and Fiber Science. – 1987. – Vol. 19. – № 4. – P. 446-452.

20. Самсонов, Ю.Н., Иванова, Г.А. Причины и последствия пожаров в бореальных лесах Сибири / Ю.Н. Самсонов, Г.А. Иванова // Регион: экономика и социология. – 2014. – №1(81). – С. 257-271.

21. Болданова, Е.В. Точка росы как основа оперативного показателя лесопожарной опасности / Е.В. Болданова // Известия ВУЗов. Лесной журнал. 2024. – №1.

22. Экспресс-картографирование повреждений лесов России пожарами по спутниковым данным Landsat / С.А. Барталев, Е.А. Лупян, Ф.В. Стыщенко [и др.] // Современные проблемы дистанционного зондирования Земли из космоса. – 2014. – Т. 11. – №1. – С. 9–20.

23. Иванов, В.А. Влияние нарушенных лесных территорий на частоту пожаров в Нижнем Приангарье / В.А. Иванов, С.А. Москальченко, Е.И. Пономарев // Хвойные бореальной зоны. – 2009. – №2 (200).

24. Спутниковый мониторинг лесных пожаров в 21 веке на территории Российской Федерации (цифры и факты по данным детектирования активного горения) / Е. А. Лупян, С. А. Барталев, И. Б. Балашов [и др.] // Современные проблемы дистанционного зондирования Земли из космоса. – 2017. – Т. 14, № 6. – С. 158-175.

25. Determination of the fire hazard of forest plantations depending on surface fuels under laboratory combustion conditions using thermophysical parameters / A.A. Malenko, A.S. Chichkarev et al. // Journal of Physics: Conference Series. – 2022. – Vol. 2233. – № 1. DOI 10.1088/1742-6596/2233/1/012010.

26. Wragg, P.D. Forbs, grasses, and grassland fire behaviour / P.D. Wragg, T. Mielke, D. Tilman // Journal of Ecology. – 2018. – Vol. 106. – № 5. – P. 1983–2001.

27. Modeling Wildfire Spread with an Irregular Graph Network / W. Jiang, F. Wang, G. Su et al. // Fire. – 2022. – № 5(6):185.

28. Волокитина, А.В., Софронова, Т.М. Картографирование растительных горючих материалов / А.В. Волокитина, Т. М. Софронова // Сибирский лесной журнал. – 2014. – № 6. – С. 8-28.

29. Спутниковое картографирование растительного покрова России по данным спектрорадиометра MODIS / С.А. Барталев, В.А. Егоров, Д.В. Ершов и др. // Современные проблемы дистанционного зондирования Земли из космоса. – 2011. – Т. 8. – № 4. – С. 285-302.

30. Плотникова, А.С. Шкала природной пожарной опасности лесных экосистем И. С. Мелехова. Обзор современных российских методических подходов / А.С. Плотникова // Вопросы лесной науки. – 2021. – Т.4. – №2.

31. Сверлова, Л. И. Метод оценки пожарной опасности в лесах по условиям погоды с учетом поясов атмосферной засушливости и сезонов года : методические рекомендации / Л. И. Сверлова ; ДальНИЦ Дальневост. УГМС. – Хабаровск, 2000. – 46 с.

32. Определение природной пожарной опасности в лесу: методические рекомендации / составители С.М. Вонский, В.А. Жданко, В.И. Корбут и др. ; Ленинградский научно-исследовательский институт лесного хозяйства (ЛенНИИЛХ). – Ленинград, 1981. – 50 с.

33. Залесов, С. В. Лесная пирология : учебник / С. В. Залесов ; Министерство науки и высшего образования Российской Федерации. –

Уральский государственный лесотехнический университет. – 4-е изд., перераб. и доп. – Екатеринбург, 2021. – 396 с.

34. Roser, M., Ortiz-Ospina, E. Global Rise of Education / M. Roser, E. Ortiz-Ospina // Our World in Data [Electronic resource]. – URL: <https://ourworldindata.org/wildfires#all-charts> (accessed: 01.06.2024).

35. Rothermel, R.C. A Mathematical model for Predicting Fire Spread in Wildland Fuels / R.C. Rothermel ; Forest Service, 1972.

36. Andrews, P.L. BehavePlus fire modeling system, version 5.0: Variables. – USDA Forest Service. Gen. Tech. Rep. Revised, 2009.

37. Баровик, Д.В., Таранчук, В.Б. Алгоритмические основы построения компьютерной модели прогноза распространения лесных пожаров // Д.В. Баровик, В.Б. Таранчук / Вестник Полоцкого государственного университета. Серия С, Фундаментальные науки: научно-теоретический журнал. 2011. – № 12. – С. 51-56.

38. Гришин, А.М. Общая математическая модель лесных пожаров и ее приложения для охраны и защиты лесов // Сопряженные задачи механики и экологии: Избранные доклады международной конференции. – 2000. – С. 88–137.

39. Баровик, Д. В. Численная реализация математической модели верховых лесных пожаров / Д. В. Баровик, В. Б. Таранчук // Весці БДПУ. Серия 3. Фізика. Математика. Інформатика. Біологія. Географія. – 2010. – № 2(64). – С. 40-44.

40. Перминов, В.А. Математическое моделирование возникновения и распространения верховых лесных пожаров в осредненной постановке // В. А. Перминов / Журнал технической физики. – 2015. – Т. 85, вып. 2. – С. 24-30.

41. Станкевич, Т.С. Прогнозирование пространственного поведения лесного пожара при неопределенности и нестационарности процесса // Известия ВУЗов. Лесной журнал. – 2021. – №1 (379).

42. Хвостиков, С.А., Баргалева, С.А. Возможности применения данных спутникового мониторинга для моделирования динамики развития природных

пожаров // Современные проблемы дистанционного зондирования Земли из космоса. – 2021. – Т. 18. – № 5. – С. 9-27.

43. Asensio, M.I. An Historical Review of the Simplified Physical Fire Spread Model PhyFire: Model and Numerical Methods / M.I. Asensio et al. // Applied Sciences. – 2023. – Vol. 13. – № 4.

44. Ghodrat, M. Software-Based Simulations of Wildfire Spread and Wind-Fire Interaction / M. Ghodrat, F. Shakeriaski, S. Fanaee, A. Simeoni // Fire. – 2023. – 6(1):12.

45. Singh, M. Forest Fire Prediction for NASA Satellite Dataset Using Machine Learning / M. Singh, C. Sharma, T. Agarwal, M.S. Pal // 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions). – Noida. – 2022. – С. 1-5.

46. Li, X., Gao, H. Prediction of Forest Fire Spread Rate Using UAV Images and an LSTM Model Considering the Interaction between Fire and Wind / X. Li, H. Gao, et al // Remote Sensing. – 2021. – 13(21):4325.

47. Natekar, S. Forest Fire Prediction using LSTM. / S. Natekar, S. Patil, A. Nair, S. Roychowdhury // 2nd International Conference for Emerging Technology. – 2021. – P. 1-5.

48. Michail D., et al. Seasonal Fire Prediction using Spatio-Temporal Deep Neural Networks. – 2024. – [Electronic resource]. – URL: <https://arxiv.org/abs/2404.06437> (дата посещения: 15.06.2024).

49. Станкевич, Т.С. Применение сверточных нейронных сетей для решения задачи оперативного прогнозирования динамики распространения лесных пожаров / Т.С. Станкевич // Бизнес-информатика. – 2018. – №4 (46).

50. Masrur, A., Manzhu, Y. Spatiotemporal attention ConvLSTM networks for predicting and physically interpreting wildfire spread / A. Masrur, Y. Manzhu // Artificial Intelligence in Earth Science. – Elsevier. – 2023. – P. 119–156.

51. Abid, F. A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems. Fire Technology / F. Abid // 2021. – Vol. 57. – P. 559–590.

52. Рыбаков, А.В., Дмитриев, А.В., Иванов, Е.В. Постановка научной задачи по обоснованию рационального размера кластера при формировании модели прогнозирования природной пожарной опасности по погодным условиям / А.В. Рыбаков, А.В. Дмитриев, Е.В. Иванов // Научные и образовательные проблемы гражданской защиты. – 2023. – №3 (58).
53. AI Journey Contest 2021: NoFireWithAI // GitHub, ai-forever/no_fire_with_ai_aij2021 [Electronic resource]. – 2021. – URL: https://github.com/ai-forever/no_fire_with_ai_aij2021 (accessed: 22.06.2024).
54. Prokhorenkova, L. CatBoost: unbiased boosting with categorical features NeurIPS / [Electronic resource]. – 2018. – URL: <https://arxiv.org/abs/1706.09516> (accessed: 14.05.2024).
55. Муньюс, С. Дж. Почасовые данные ERA5-Land с 1950 года по настоящее время. Служба по изменению климата Copernicus (C3S): [Хранилище климатических данных (CDS)]. – 2019. – DOI: 10.24381/cds.e2161bac (дата доступа: 15.04.2024).
56. Титаев, П.С. Имитационная модель вероятность возникновения лесных пожаров с учетом // Пожаровзрывобезопасность. – 2015. – Т. 24. – №5. – С. 44-53.
57. OpenStreetMap [Electronic resource]. – URL: opendatacommons.org (accessed: 14.06.2024).
58. Elsayed, S. et al. Do We Really Need Deep Learning Models for Time Series Forecasting? [Electronic resource]. – 2021. – URL: <https://arxiv.org/abs/2101.02118> (accessed: 14.06.2024).
59. Ribeiro, M., Coelho, L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series / M. Ribeiro, L. Coelho // Applied Soft Computing. – 2020. – V. 86.
60. Fire Information for Resource Management System / NASA-FIRMS // [Electronic resource]. – URL: <https://firms.modaps.eosdis.nasa.gov/map/> (accessed: 22.06.2024).

61. Крюкова, С.В., Симакина, Т.Е. Оценка методов пространственной интерполяции метеорологических данных / С.В. Крюкова, Т.Е. Симакина // Общество. Среда. Развитие (Terra Humana). – 2018. – №1 (46).

62. Ярмоленко, А.С., Куцаева, О.А. Обоснование метода радиальных базисных функций для классификации геопространственных объектов / А.С. Ярмоленко, О.А. Куцаева // Вестник Полоцкого государственного университета. Серия F. Строительство. Прикладные науки. – 2017. – №8.

63. Скворцов, А.В. Триангуляция Делоне и её применение. – Томск: Изд-во Том. ун-та. – 2002. – 128 с.