

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ/SYSTEM ANALYSIS,  
MANAGEMENT AND PROCESSING OF INFORMATIONDOI: <https://doi.org/10.60797/IRJ.2025.156.88>ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРЕДВАРИТЕЛЬНОГО АНАЛИЗА ДАННЫХ  
О ЛЕСНЫХ ПОЖАРАХ РЕСПУБЛИКИ БУРЯТИЯ

Научная статья

Бильгаева Л.П.<sup>1,\*</sup>, Евдокимова И.С.<sup>2</sup>, Главинская А.А.<sup>3</sup><sup>1</sup>ORCID : 0000-0002-8991-3206;<sup>1</sup> Федеральное государственное бюджетное образовательное учреждение высшего образования «Восточно-Сибирский государственный университет технологий и управления», Улан-Удэ, Российская Федерация<sup>2,3</sup> Федеральное государственное бюджетное образовательное учреждение высшего образования «Восточно-Сибирский государственный университет технологий и управления», Улан-Удэ, Российская Федерация

\* Корреспондирующий автор (bilgaeval[at]mail.ru)

**Аннотация**

Статья посвящена предварительному анализу данных о лесных пожарах в Республике Бурятия на основе информации Геоинформационного портала «Природа Бурятии». В процессе проведенных исследований было построено пространственное и временное распределение лесных пожаров. Приведена динамика пожаров за период с 2015 по 2024 гг. и структура причин возгораний, анализ которой показал преобладание антропогенного фактора и природного в виде гроз. Концентрация очагов возгораний и интенсивность пожаров исследовалась на основе построения тепловой карты. Анализ информативности имеющихся данных и их визуализация проводились с помощью методов машинного обучения. Анализ данных проводился с помощью программного обеспечения, разработанного на языке программирования Python с использованием библиотек Pandas, Scikit-Learn, Matplotlib, Folium.

**Ключевые слова:** лесные пожары, машинное обучение, метод k-средних, Евклидово расстояние, метод локтя, визуализация, тепловая карта.

APPLICATION OF MACHINE LEARNING METHODS FOR PRELIMINARY ANALYSES OF FOREST FIRES  
DATA IN THE REPUBLIC OF BURYATIA

Research article

Bilgaeva L.P.<sup>1,\*</sup>, Yevdokimova I.S.<sup>2</sup>, Glavinskaya A.A.<sup>3</sup><sup>1</sup>ORCID : 0000-0002-8991-3206;<sup>1</sup> Federal State Budgetary Educational Institution of Higher Education "East Siberian State University of Technology and Management", Ulan-Ude, Russian Federation<sup>2,3</sup> East Siberia State University of Technology and Management, Ulan-Ude, Russian Federation

\* Corresponding author (bilgaeval[at]mail.ru)

**Abstract**

The article is devoted to a preliminary analysis of data on forest fires in the Republic of Buryatia based on information from the Geoinformation Portal 'Nature of Buryatia'. In the process of research, the spatial and temporal distribution of forest fires was constructed. The dynamics of fires for the period from 2015 to 2024 was provided, as well as the structure of the causes of fires, the analysis of which showed the predominance of anthropogenic natural factors in the form of thunderstorms. The concentration of fire centres and the intensity of fires were studied based on the construction of a heat map. Analysis of the informativity of the available data and their visualisation was carried out using machine learning methods. The data were analysed using software developed in the Python programming language using Pandas, Scikit-Learn, Matplotlib, Folium libraries.

**Keywords:** forest fires, machine learning, k-means method, Euclidean distance, elbow method, visualisation, heat map.

**Введение**

Пожары представляют собой одну из самых серьезных угроз для экосистем, человеческой жизни и материальных ценностей. В последние десятилетия в Бурятии наблюдается рост числа и интенсивности лесных пожаров, что подчеркивает необходимость глубокого анализа данных о пожарах для разработки эффективных стратегий их предотвращения и реагирования. С учетом изменения климата, увеличения урбанизации и человеческой деятельности, понимание причин и последствий пожаров становится более актуальным, чем когда-либо.

Анализ данных о пожарах включает в себя сбор, обработку и интерпретацию информации, связанной с различными аспектами возникновения и распространения огня. Это может включать данные о погодных условиях, типах растительности, географическом положении, а также информацию о предыдущих инцидентах. С помощью современных методов анализа данных, таких как машинное обучение и статистические модели, можно выявлять основные паттерны и аномалии, которые могут помочь в понимании причин и последствий пожаров, а также предсказать вероятность возникновения пожара и его потенциальную интенсивность.

В [1] выполнен анализ динамики лесных пожаров в Республике Бурятия за период с 2013 по 2019 год, включая их причины, последствия и уязвимость территории. Основное внимание уделено антропогенным факторам

возникновения пожаров, а также влиянию климатических условий и видового состава лесов. Анализ проведен с использованием программного средства Microsoft Excel.

Другое направление исследования пожаров посвящено влиянию дорожной инфраструктуры на возникновение, распространение и тушение лесных пожаров в Восточно-Сибирской тайге, используя комбинацию данных дистанционного зондирования и наблюдений местных сообществ [2].

В статье [3] применяется кластерный анализ для обработки статистических данных о лесных пожарах вблизи железнодорожной инфраструктуры на примере Республики Бурятия. Авторы исследуют взаимосвязь между антропогенной нагрузкой, вызванной железнодорожными объектами, и возникновением пожаров, используя метеорологические данные, характеристики лесных массивов и пространственное распределение пожаров.

В работе [4] проанализированы данные Республики Бурятия: выявляются статистические связи между погодными факторами (наземными и спутниковыми измерениями, включая влажность почвы) и статистикой лесных пожаров. Авторы показывают, как сочетание метеорологических параметров коррелирует с возникновением пожаров, что помогает оценивать пожарную опасность на основе локальных условий в Бурятии.

Исследование посвящено региону Юго-Восточной Сибири (включая Байкальский/Забайкальский регионы, сходные по условиям с Бурятией). Авторы [5] разработали модель глубокого обучения для выявления влияния климатических аномалий на пожары и прогнозирования их интенсивности. Модель выявила, например, что положительная фаза североатлантической тропической осцилляции весной усиливает сушь и таяние снега, что приводит к резкому росту площади весенних пожаров в регионе. Прогнозы показывают возможное усиление экстремальных пожаров в будущем при изменении климата.

В [6] предложен комплексный подход к прогнозированию возникновения лесных пожаров, объединяющий разнообразные данные (метеоусловия, Геопространственные факторы, антропогенные показатели) и современные алгоритмы машинного обучения. Авторы разработали единую конвейерную систему сбора и предобработки данных и протестировали несколько моделей — случайный лес, градиентный бустинг (XGBoost), автоэнкодер для поиска аномалий, ConvLSTM для временных рядов, Attention-MLP и сверточную сеть RegNetX.

Целью данной статьи является предоставление результатов предварительного анализа данных о пожарах в Республике Бурятия путем исследования ключевых факторов, способствующих их возникновению и распространению, с использованием исторических данных Геоинформационного портала «Природа Бурятии» [7] за период с 2015 по 2024 гг. и современных методов анализа таких, как методы машинного обучения и визуализации данных.

В процессе предварительного анализа исходных данных были выявлены проблемы в их структуре: несогласованность названий признаков; существенное сокращение признаков в данных за 2021–2022 гг.; сокращённые заголовки в данных за 2021 и 2023 гг.

Такой подход демонстрирует важность предварительной обработки исходных данных и их визуализации как инструмента для улучшения понимания динамики пожаров и повышения готовности к ним. А также являются основой для дальнейших исследований.

### Методы и принципы исследования

Для предварительного анализа данных о пожарах Республики Бурятия использовался метод кластеризации k-means [8]. Этот метод относится к технологии машинного обучения. Он использовался для группировки пожаров по схожим признакам. В качестве метрики расстояния между центроидами использовалась метрика «Евклидово расстояние», вычисляемое по формуле

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

где  $p$  и  $q$  — координаты двух точек на плоскости.

Оптимизация количества кластеров выполнялась по методу локтя [9] с применением метрики SSE (сумма квадратов расстояний), которая измеряет, насколько близко точки находятся к центрам своих кластеров. Суть данного метода заключается в том, чтобы найти точку, в которой добавление дополнительных кластеров перестает значительно уменьшать сумму квадратов расстояний между точками и центрами кластеров.

Для визуализации исходных данных в задаче предварительного анализа данных о пожарах были построены диаграммы по различным признакам. Также была построена тепловая карта пожаров, которая позволяет быстро и наглядно оценить распределение и интенсивность пожаров по районам Республики Бурятия.

Для анализа и визуализации исходных данных о пожарах было разработано программное средство на языке программирования Python с использованием таких библиотек, как Pandas, Scikit-Learn, Matplotlib, Folium [10].

### Результаты анализа

В течение последних 10 лет (с 2015 по 2024 гг.) наблюдается нестабильная динамика количества пожаров: с 2015 по 2021 год отмечается снижение, однако в 2023–2024 годах происходит новый рост (рис. 1).



Рисунок 1 - Динамика пожаров с 2015 по 2024 гг  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.1>

На рисунке 2 представлена диаграмма, отражающая структуру причин пожаров.

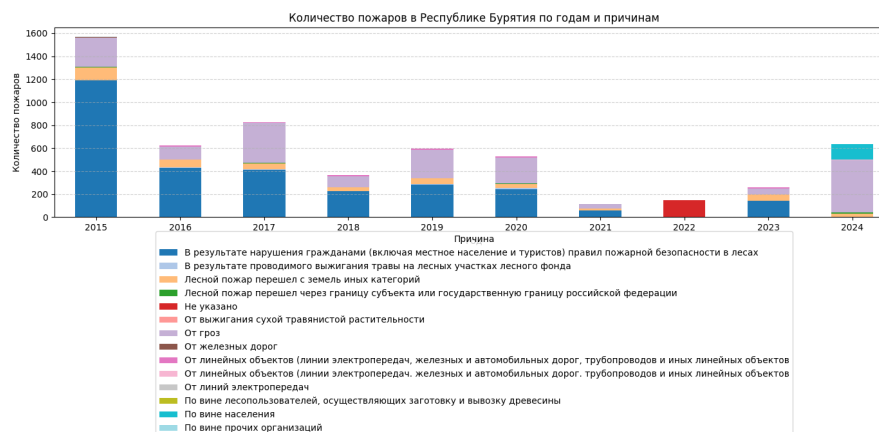


Рисунок 2 - Структура причин пожаров  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.2>

Анализ диаграммы показывает, что структура причин пожаров за этот период со временем становится более разнообразной, однако человеческий фактор по-прежнему преобладает. Рост количества пожаров в 2024 году частично связан с естественными причинами — в частности, грозами, что может быть следствием изменения климата.

На рисунке 3 представлена диаграмма, в которой иллюстрируются способы обнаружения пожаров. Анализ этой диаграммы показывает, что до 2021 года большая часть данных по способам обнаружения не указывалась. Этот факт ограничивает процесс анализа. Однако из имеющихся данных можно выделить такие способы обнаружения пожаров по убыванию их количества, как «Дистанционное зондирование Земли», «Лесоавиационные работы» и «Наземное патрулирование».

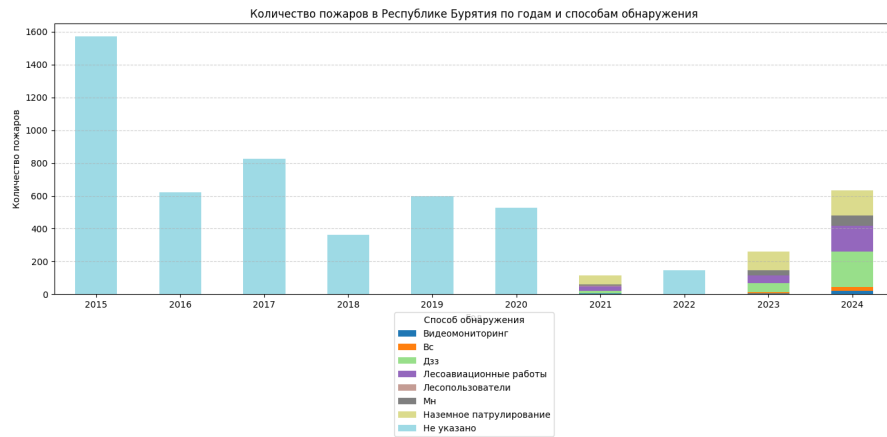


Рисунок 3 - Способы обнаружения пожаров  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.3>

Пространственный анализ данных показал, что наибольшее количество пожаров зафиксировано в Заиграевском, Прибайкальском и Северо-Байкальском районах. Это может быть связано с такими факторами, как климатические условия, плотность населения и особенности ландшафта. (рис. 4). В целом диаграмма демонстрирует значительные различия в количестве пожаров между районами республики.

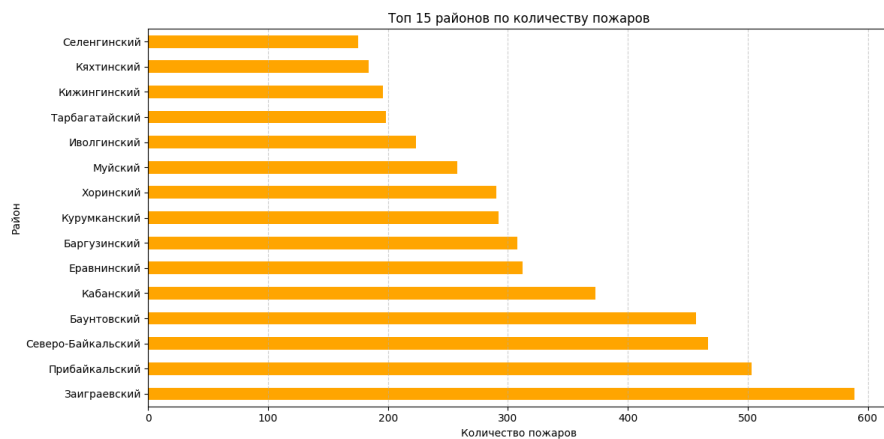


Рисунок 4 - Пространственный анализ пожаров  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.4>

Анализ сезонности показывает, что пик пожаров приходится на летние месяцы, особенно на июнь, как показано на рисунке 5. Это может быть связано с высокой температурой, засушливой погодой и активной человеческой деятельностью в этот период.



Рисунок 5 - Распределение пожаров по месяцам  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.5>

Наибольшее число возгораний происходит в дневное время, особенно после полудня, с пиком около 15:00 часов. В этот период времени наблюдается наибольшая активность населения и использование различного оборудования и техники. В ночное время количество пожаров резко снижается, как это можно видеть на рисунке 6.

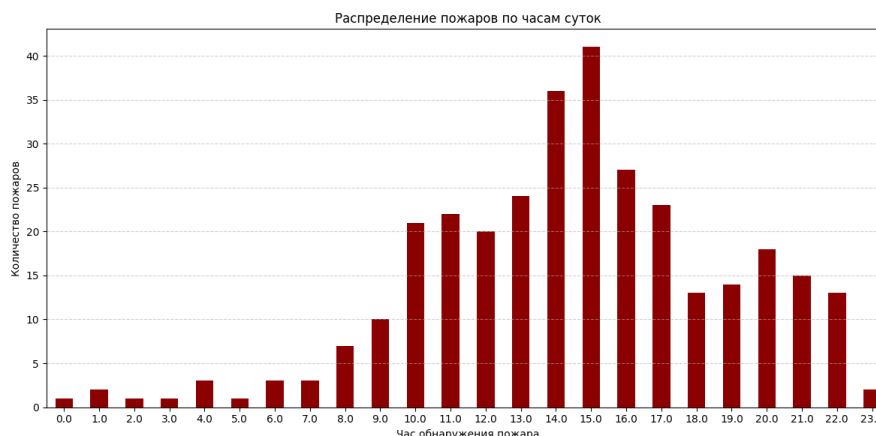


Рисунок 6 - Распределение пожаров по часам суток  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.6>

Низкая активность в ночное время может быть обусловлена снижением человеческой активности и более низкими температурами.

Частота пожаров возрастает в конце рабочей недели и в выходные дни, особенно в пятницу. Это может быть связано с выездами населения на природу и снижением уровня соблюдения мер пожарной безопасности (рис. 7).

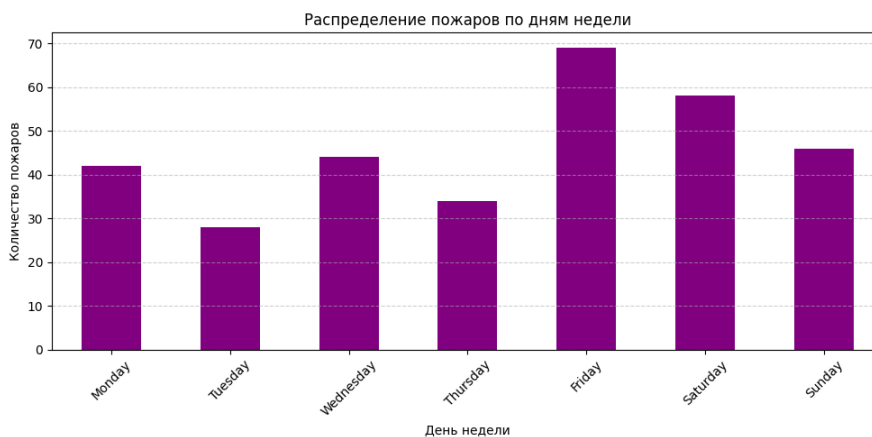


Рисунок 7 - Распределение пожаров по дням недели  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.7>

По районам республики также наблюдаются различия в распределении пожаров по дням недели. Так, Северо-Байкальский район показывает пик активности в воскресенье, а Баунтовский район — в начале недели. Другие районы, такие как Прибайкальский, Кабанский и Заиграевский, демонстрируют относительно низкие показатели пожаров на протяжении всей недели (рис. 8).

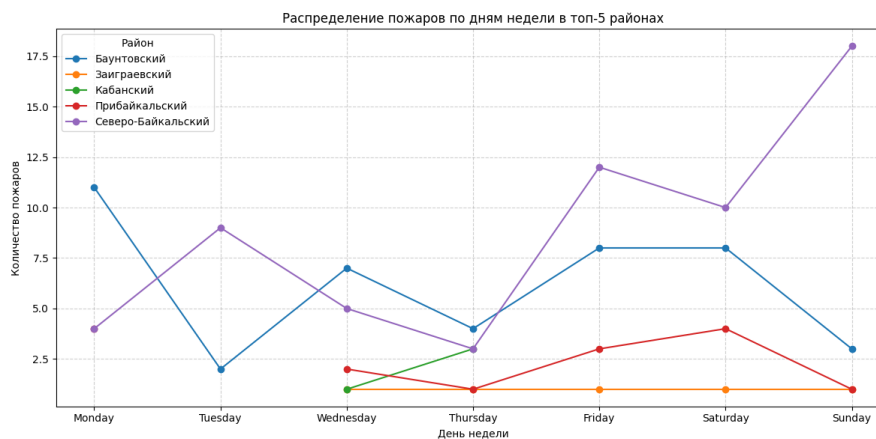


Рисунок 8 - Распределение пожаров по дням недели в топ-5 районах  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.8>

На тепловой карте, построенной по данным 2015–2024 годов, представленной на рисунке 9, видно, что наибольшая плотность пожаров наблюдается в центральной и северо-восточной частях республики.

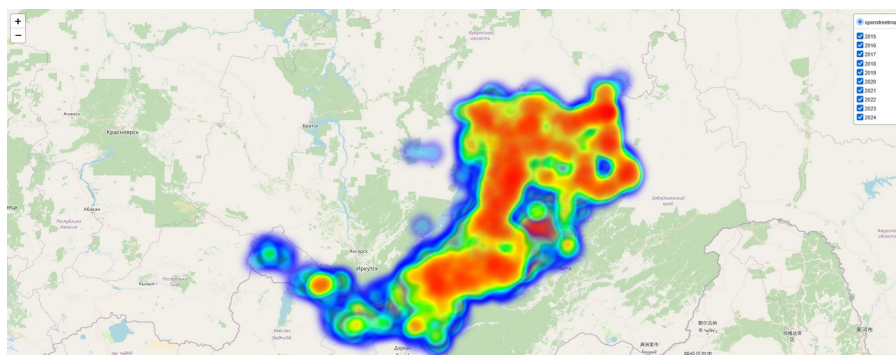


Рисунок 9 - Тепловая карта пожаров по данным 2015-2024 гг  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.9>

Красные и жёлтые зоны на карте указывают на концентрацию очагов возгорания. Зелёный и синий цвета обозначают области с меньшей плотностью пожаров. Преобладание красных и желтых зон на тепловой карте

свидетельствуют о наиболее уязвимых площадях республики Бурятия с точки зрения возгораний возможно из-за климатических условий. Здесь требуется более детальный анализ, который планируется провести в будущем путем построения карты рисков с использованием пространственной и временной информации.

Был проведён кластерный анализ с использованием 7 признаков, выделенных из исходных данных, которые представлены в таблице 1. Всего было определено 5 кластеров с количеством пожаров в каждом из них.

Анализ кластеров показал, что такие признаки, как: «Категория земель» (преобладает значение «Естественные насаждения»); «Тип покрова» (доминирует «Травяной»); «Характер» (в основном «Хвойные насаждения»); «Порода» (большое количество схожих значений) — имеют низкую информативность и не способствуют выделению уникальных кластеров. Отсюда следует вывод, что не следует их использовать в дальнейшем анализе.

На рисунке 10 представлена карта, отображающая выделенные первичные кластеры по всем 7 признакам.

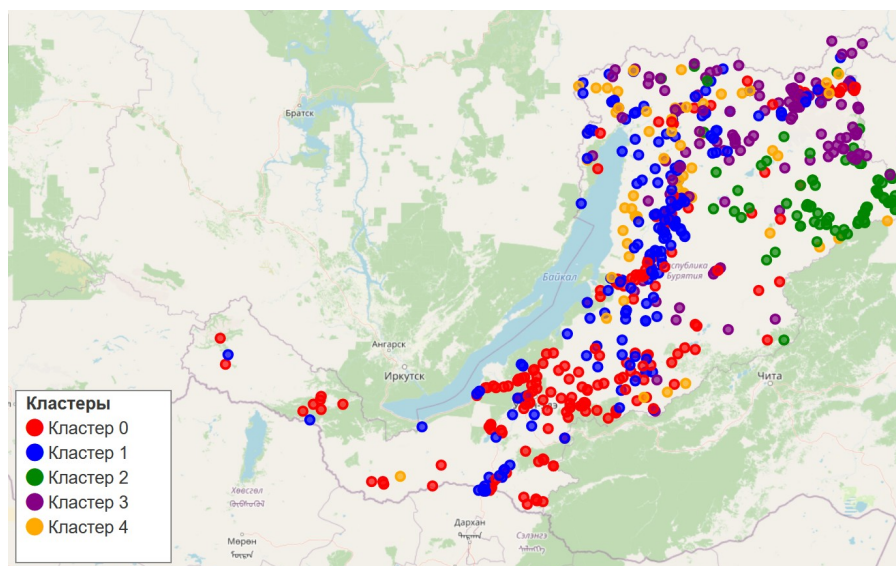


Рисунок 10 - Результаты кластерного анализа  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.10>

Таблица 1 - Характеристики кластеров  
DOI: <https://doi.org/10.60797/IRJ.2025.156.88.11>

Характеристики	Кластер 0	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Количество пожаров	206	168	78	122	60
Причина	По вине населения	От гроз	От гроз	От гроз	От гроз
%	58,8	98,2	87,2	92,6	91,7
Способ обнаружения	Наземное патрулирование	Лесоавиационные работы	Дистанционное зондирование Земли	Дистанционное зондирование Земли	Лесоавиационные работы
%	56,3	58,9	100	77,9	43,3
Тип покрова	Травяной	Травяной	Травяной	Травяной	Лишайники
%	93,7	97,6	80,8	77,9	51,7
Категория лесов	Эксплуатационный	Защитный	Резерв	Эксплуатационный	Защитный
%	52,4	71,4	80,8	74,6	61,7
Категория земель	Естественные насаждения	Естественные насаждения	Естественные насаждения	Естественные насаждения	Естественные насаждения
%	96,1	100	82,2	98,4	85,0
Характер	Хвойный	Хвойный	Лиственный	Хвойный	Хвойный
%	85,4	96,6	78,2	95,9	76,6



Характеристики	Кластер 0	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Порода	Сосна	Сосна	Кустарники	Лиственница	Кедровый стланик
%	74,8	68,5	-	60,3	31,7

После исключения признаков с низкой информативностью был проведён повторный кластерный анализ, в результате которого определены 5 кластеров:

- кластер 0 — пожары, вызванные антропогенным фактором;
- кластеры 1–3 — пожары, возникшие вследствие гроз;
- кластер 4 — пожары с иных земель.

В таблице 2 представлены уточненные характеристики кластеров такие, как «Количество пожаров», «Способ обнаружения», «Категория лесов»

На карте рисунка 11 представлены обновлённые кластеры, более точно отражающие типы и природу пожаров.

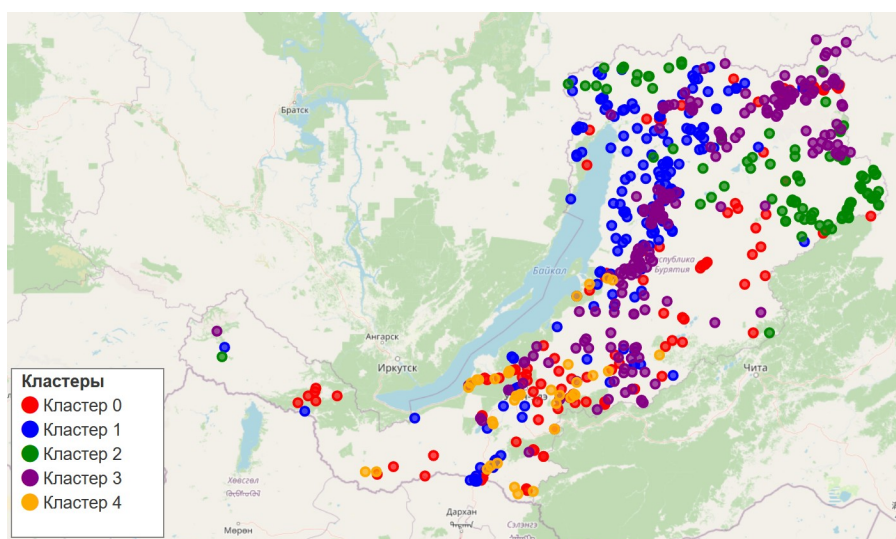


Рисунок 11 - Кластеры с уточненными характеристиками

DOI: <https://doi.org/10.60797/IRJ.2025.156.88.12>

Пожары, вызванные антропогенным фактором, сосредоточены вблизи населенных пунктов или вблизи автомобильных дорог. А пожары, вызванные природным фактором, а именно вследствие гроз, как правило сосредоточены в гористой местности и в тайге. Для более точного анализа этих факторов необходимо провести более детальные исследования с использованием космоснимков.

Таблица 2 - Уточненные характеристики кластеров

DOI: <https://doi.org/10.60797/IRJ.2025.156.88.13>

Характеристики	Кластер 0	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Количество пожаров	127	173	95	201	38
Причина	По вине населения	От гроз	От гроз	От гроз	Пожар с иных земель
%	98,4	100	88,4	98,0	84,2
Способ обнаружения	Наземное патрулирование	Лесоавиационные работы	Дистанционное зондирование Земли	Лесоавиационные работы	Наземное патрулирование
%	44,9	36,4	96,8	33,8	63,2
Категория лесов	Защитный	Защитный	Резерв	Эксплуатационный	Защитный
%	50,4	98,3	93,7	99,5	89,5



## Обсуждение

Данное исследование выполнялось в рамках республиканского гранта «Развитие методов предупреждения лесных пожаров на территории Республики Бурятия». Была поставлена конкретная цель предварительного анализа данных Геоинформационного портала «Природа Бурятии», на основе которой были определены и решаются задачи для более глубокого анализа ситуации о лесных пожарах в Республике Бурятия, в том числе и препроцессинг исходных данных.

Проведенный предварительный анализ лесных пожаров в Республике Бурятия по историческим данным Геоинформационного портала «Природа Бурятии» показал, что наиболее частые очаги возгорания лесных массивов, находящихся вблизи населенных пунктов, происходят по причине антропогенных факторов. Природные факторы, такие как грозы, являются причиной возгорания в лесных массивах, удаленных от населенных пунктов, и горных местностях. Наиболее частые пожары возникают в определенные временные интервалы: это начало летних месяцев (июнь), когда население активизируется в связи с началом сельскохозяйственных работ, во время отпусков, по дням недели — это конец рабочей недели, когда население выезжает отдыхать на природу, а также по времени суток — это дневные часы. Распределение пожаров по районам республики показывает наиболее уязвимые районы.

Тепловая карта демонстрирует интенсивность и частоту возгораний в центральной и северо-восточной частях республики. Для более глубокого анализа причин возгорания и точного определения конкретных мест необходимо проведение дополнительных исследований с использованием спутниковых снимков для применения эффективных методов машинного обучения и графовых нейронных сетей.

При решении задачи кластеризации методом k-Means были использованы данные за 2024 без отсутствующих значений. При реализации алгоритма k-Means для инициализации центроидов был использован метод k-Meanss++, при этом первый центроид выбирался случайным образом, а каждый последующий посредством метрики «Евклидово расстояние». Всего было выполнено 3000 итераций. Оптимальное количество кластеров определено с помощью метода локтя.

Кластеризация исходных данных позволила выявить основные признаки, по которым нужно проводить дальнейший анализ. Получено пять кластеров, демонстрирующие группировку данных по семи признакам, анализ которых показывает, что информативность некоторых признаков таких, как «Категория земель» и «Тип покрова» является низкой и не способствует выделению уникальных кластеров по этим признакам. Это связано с тем, что на территории Республики Бурятия доминирует категория земель «Естественные насаждения» и тип покрова «Травяной». Поэтому была проведена повторная кластеризация без низкоинформативных признаков, и в результате было получено также пять кластеров.

Обзор литературных источников позволил ознакомиться с текущей ситуацией исследований в области анализа лесных пожаров на территории Республики Бурятия и других близ лежащих к ней регионов, а также определиться с методами дальнейшего исследования по анализу и предсказания лесных пожаров в Республике Бурятия.

## Заключение

Анализ данных о пожарах в Республике Бурятия за период с 2015 по 2024 гг. показал, что:

- наблюдается колебание числа пожаров с пиком в 2015 г. и повторным ростом в 2023–2024 гг.;
- пожары чаще происходят летом, днём и в выходные дни;
- основными причинами остаются антропогенные факторы и грозы;
- пространственное распределение пожаров указывает на то, что они сосредоточены в центральной и северо-восточной части республики;
- кластеризация позволила выделить 5 типов пожаров с разной природой возникновения.

В дальнейшем полученные результаты будут использованы для построения карты рисков с расчетом вероятности возникновения лесных пожаров на территории Республики Бурятия с учетом текущих пространственных и временных данных на основе машинного обучения и графовых нейронных сетей.

## Конфликт интересов

Не указан.

## Conflict of Interest

None declared.

## Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

## Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

## Список литературы / References

1. Игнатъева А.В. Динамика лесных пожаров в республике Бурятия / А.В. Игнатъева, Н.В. Барановский. — URL: <https://jfsi.ru/5-2-2022-ignatvabarankovskiy/> (дата обращения: 21.03.2025).
2. Kuklina V. Combining community observations and remote sensing to examine the effects of roads on wildfires in the East Siberian boreal forest / V. Kuklina, O. Sizov, V. Bogdanov [et al.] // Arctic Science. — 2023. — № 9 (2). — P. 393–407. — DOI: 10.1139/as-2021-0042.
3. Belikova M. Analysis of Spatial Distribution Processes for Forest Fires Near the Railway Infrastructure Using Clustering: Case Study / M. Belikova, N. Baranovskiy, A. Karanin [et al.] // International Journal on Engineering Applications (IREA). — 2022. — № 10. — P. 411. — DOI: 10.15866/irea.v10i6.22224.

4. Badmaev N.B. Correlation analysis of terrestrial and satellite meteorological data in the territory of the Republic of Buryatia (Eastern Siberia, Russian Federation) with forest fire statistics / N.B. Badmaev, A. Bazarov // *Agricultural and Forest Meteorology*. — 2020. — № 297. — P. 108245. — DOI: 10.1016/j.agrformet.2020.108245.
5. Gui K. Future climate-driven escalation of Southeastern Siberia wildfires revealed by deep learning / K. Gui, X. Zhang, H. Che [et al.] // *npj Clim Atmos Sci*. — 2024. — № 7. — P. 263. — DOI: 10.1038/s41612-024-00815-x.
6. Illarionova S. Exploration of geo-spatial data and machine learning algorithms for robust wildfire occurrence prediction / S. Illarionova, D. Shadrin, F. Gubanov [et al.] // *Sci Rep*. — 2025. — № 15. — Article 10712. — DOI: 10.1038/s41598-025-94002-4.
7. Геоинформационный портал «Природа Бурятии». — URL: [https://prioda-rb.ru](https://priroda-rb.ru) (дата обращения: 21.03.2025).
8. Aggarwal C.C. *Data Mining. The Textbook* / C.C. Aggarwa. — New York: Springer, 2015. — P. 162–163.
9. Umargono E. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula / E. Umargono, J. Suseno, S.K. Gunawan. — 2020. — DOI: 10.2991/assehr.k.201010.019.
10. Маккинни У. *Python и анализ данных* / У. Маккинни. — Москва: ДМК Пресс, 2015. — 482 с.

### Список литературы на английском языке / References in English

1. Ignat'eva A.V. *Dinamika lesnyh pozharov v respublike Burjatija* [Dynamics of forest fires in the Republic of Buryatia] / A.V. Ignat'eva, N.V. Baranovskij. — URL: <https://jfsi.ru/5-2-2022-ignatevabaranovskiy/> (accessed: 21.03.2025). [in Russian]
2. Kuklina V. Combining community observations and remote sensing to examine the effects of roads on wildfires in the East Siberian boreal forest / V. Kuklina, O. Sizov, V. Bogdanov [et al.] // *Arctic Science*. — 2023. — № 9 (2). — P. 393–407. — DOI: 10.1139/as-2021-0042.
3. Belikova M. Analysis of Spatial Distribution Processes for Forest Fires Near the Railway Infrastructure Using Clustering: Case Study / M. Belikova, N. Baranovskiy, A. Karanin [et al.] // *International Journal on Engineering Applications (IREA)*. — 2022. — № 10. — P. 411. — DOI: 10.15866/irea.v10i6.22224.
4. Badmaev N.B. Correlation analysis of terrestrial and satellite meteorological data in the territory of the Republic of Buryatia (Eastern Siberia, Russian Federation) with forest fire statistics / N.B. Badmaev, A. Bazarov // *Agricultural and Forest Meteorology*. — 2020. — № 297. — P. 108245. — DOI: 10.1016/j.agrformet.2020.108245.
5. Gui K. Future climate-driven escalation of Southeastern Siberia wildfires revealed by deep learning / K. Gui, X. Zhang, H. Che [et al.] // *npj Clim Atmos Sci*. — 2024. — № 7. — P. 263. — DOI: 10.1038/s41612-024-00815-x.
6. Illarionova S. Exploration of geo-spatial data and machine learning algorithms for robust wildfire occurrence prediction / S. Illarionova, D. Shadrin, F. Gubanov [et al.] // *Sci Rep*. — 2025. — № 15. — Article 10712. — DOI: 10.1038/s41598-025-94002-4.
7. Geoinformacionnyj portal «Priroda Burjatii» [Geoinformation portal 'Nature of Buryatia']. — URL: [https://prioda-rb.ru](https://priroda-rb.ru) (accessed: 21.03.2025). [in Russian]
8. Aggarwal C.C. *Data Mining. The Textbook* / C.C. Aggarwa. — New York: Springer, 2015. — P. 162–163.
9. Umargono E. K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula / E. Umargono, J. Suseno, S.K. Gunawan. — 2020. — DOI: 10.2991/assehr.k.201010.019.
10. McKinney W. *Python i analiz dannyh* [Python and data analysis] / W. McKinney. — Moskva: DMK Press, 2015. — 482 p. [in Russian]