

Исследование влияния грозовых явлений на лесные пожары на основе методов машинного обучения

Юлия Пестова^{1,2}, Ольга Николайчук^{1,2}, Яна Ветвицкая¹

¹ Иркутский государственный университет, Россия, 66400, г. Иркутск, ул. Карла Маркса 1

² Институт динамики систем и теории управления им. В.М. Матросова Сибирского отделения Российской академии наук, Россия, 664033, г. Иркутск, ул. Лермонтова, 134

Аннотация

Исследование посвящено анализу влияния грозовых явлений на риск возникновения лесных пожаров с использованием методов машинного обучения. Проведено сопоставление данных о грозах и лесных пожарах в Иркутской области за период с 2011 по 2023 год с применением алгоритма ближайшего соседа KD Tree. Показано, что включение грозовых признаков незначительно улучшает точность прогнозирования риска, однако ведущую роль играют погодные условия, предшествующие грозе. Полученные результаты могут быть использованы для совершенствования моделей прогнозирования и предупреждения лесных пожаров.

Ключевые слова

Риск лесных пожаров, сухие грозы, грозовые пожары, машинное обучение, прогнозирование пожаров, KD Tree, AutoGluon

1. Введение

Лесные пожары остаются серьёзной экологической проблемой, представляющей значительный риск для природных экосистем и хозяйственной деятельности человека. Среди причин возникновения пожаров, особую роль играют молниевые разряды, вызывающие до 10% лесных пожаров в России [1]. С ростом температуры и изменением циркуляции атмосферы, согласно прогнозам, количество грозовых явлений и связанных с ними пожаров в ближайшие годы будет расти [2, 3]. Целью исследования является оценка влияния грозовых признаков на точность модели прогнозирования риска возникновения лесных пожаров с помощью машинного обучения.

2. Методология исследования и предварительная обработка данных

В исследовании собраны и обработаны данные о лесных пожарах и грозах в Иркутской области за период с 2011 по 2023 год. Исходный набор данных для исследования содержит записи о грозах, полученных из архива NASA GHRC (WWLLN) [4], о лесных пожарах, предоставленных ИКИ РАН [5] и физико-географических факторов территории, включая: топографические (высота, крутизна и ориентация наклона), характеристики растительного покрова [6] и метеорологические факторы [7, 8].

На этапе обработки данных проведена пространственно-временная фильтрация данных о пожарах, генерация «не пожарных» событий в равном соотношении с образцами «пожарных» событий, и их агрегирование с метеорологическими показателями за периоды 3, 7 и 14 дней. В результате был получен набор данных о событиях с учетом пространственных характеристик территории и метеорологических показателей, таких как сумма осадков, максимальные и

7th International Workshop on Information, Computation, and Control Systems for Distributed Environments (ICCS-DE 2025), July 7–11, 2025, Irkutsk, Russia

EMAIL: yupest@icc.ru (A. 1); nikoly@icc.ru (A. 2); yanavetvitskaya@gmail.com (A. 3)

ORCID: 0000-0001-5874-7965 (A. 1); 0000-0002-5186-0073 (A. 2); 0009-0008-9546-5586 (A. 3)



©2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

ICCS-DE 2025 Workshop Proceedings

DOI: 10.47350/ICCS-DE.2025.13

минимальные температуры воздуха, средние показатели скорости ветра, влажности воздуха, почвы и количество засушливых дней.

Для пространственно-временного сопоставления событий пожаров и гроз использовался алгоритм ближайшего соседа KD Tree. KD Tree представляет собой структуру данных, основанную на многомерном бинарном дереве, эффективно работающую при поиске ближайших соседей в пространствах с умеренным числом измерений и при использовании стандартных евклидовых метрик. В результате проведённых экспериментов, алгоритм KD Tree продемонстрировал высокую скорость работы и точность для решения задачи сопоставления событий пожаров и гроз в трёхмерном пространстве, на основе географических координат и даты событий.

В данном исследовании гроза считается причиной пожара, если выполняются условия: наличие сухой грозы (при осадках менее, чем 0,5 мм), дистанция до пожара менее 15 км или принадлежность к одной ячейке регулярной сетки 0,25°, временной период между грозой и пожаром менее 7 дней и наличие как минимум 3-х засушливых дней до пожара.

Различные исследования подтверждают, что большинство пожаров, вызванных молниями, возникает в течение первых 2–4 дней после грозового разряда и на расстоянии до 20 км [2]. В данной работе также было установлено, что большинство пожаров возникают в первые 3 дня после грозы, особенно часто — непосредственно в день грозы. Наибольшее количество пожаров от грозы возникает в пределах 5–10 км. Согласно статье [2], максимальная вероятность возгорания наблюдается при условии, если грозе предшествовал период без значительных осадков, что подтверждает важность учёта засушливых дней перед грозой при прогнозировании риска пожаров.

Результаты анализа показали, что в среднем число грозовых пожаров составляет до 6% от общего числа возгораний, в определенные годы этот процент увеличивается до 8-9%. Рисунок отражает статистику грозовых пожаров на территории Иркутской области за 2013 г. (рис. 1).

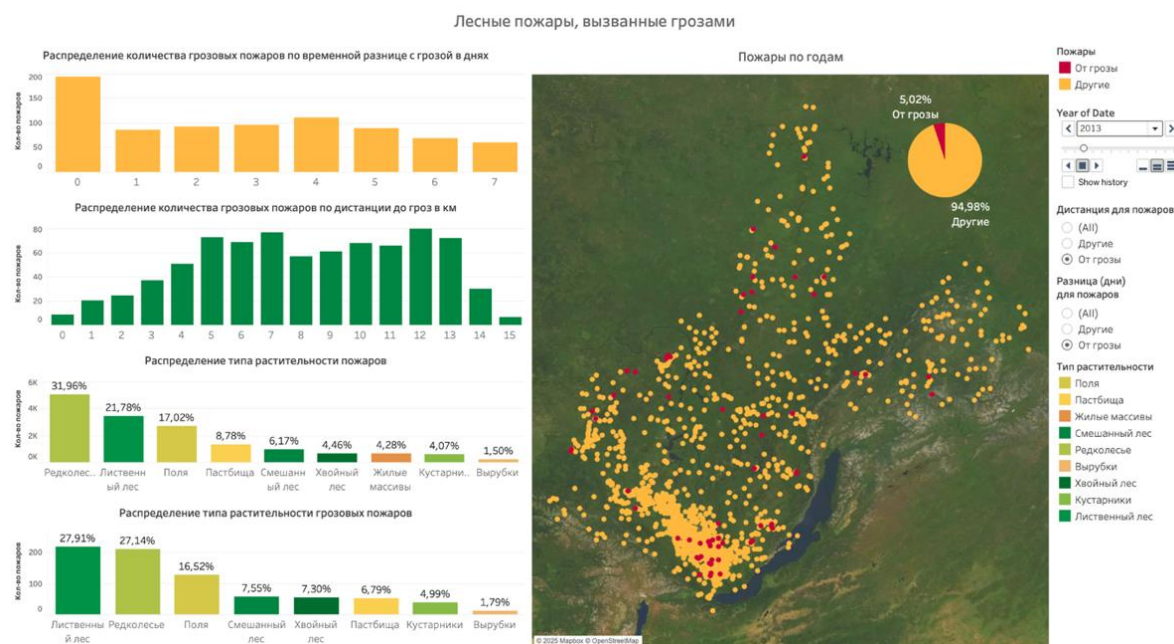


Рисунок 1: Визуализации распределений и картирование пожаров, в том числе от гроз

Исследование пространственного распределения сухих гроз по территории Иркутской области показал, что молниевые разряды распределяются неравномерно и сконцентрированы преимущественно в центральной и южной частях региона. Максимальная плотность сухих гроз наблюдается в районе, прилегающем к озеру Байкал, а также вдоль рек Ангара и Лена, что обусловлено особенностями местного рельефа и микроклимата этих территорий. Рельеф, в частности, наличие возвышенностей и горных систем, таких как Восточный Саян и Прибайкалье, способствует формированию конвективных облаков и, следовательно,

увеличивает вероятность сухих гроз и связанных с ними пожаров. В северной части области частота сухих гроз заметно снижается, что связано с климатическими особенностями этой территории – более прохладными и влажными условиями, менее способствующими развитию засух и образованию сухих гроз. Рисунок позволяет ознакомиться с картированием сухих гроз (рис. 2).



Рисунок 2: Картирование сухих гроз за 2014 год

3. Результаты моделирования и анализ важности признаков

Для построения моделей классификации риска возникновения лесных пожаров была использована логистическая регрессия как базовая модель. Точность данной модели составила 0,92. Затем была применена библиотека автоматического машинного обучения AutoGluon, позволяющая автоматически подбирать гиперпараметры моделей машинного обучения, обучать их, в том числе ансамбли и деревья решений, и выявлять наиболее производительную. В процессе автоматического подбора моделей AutoGluon использовались модели Random Forest, LightGBM, ExtraTrees, CatBoost, где наиболее эффективной оказалась модель случайного леса (Random Forest [9]), обученная на метеорологических показателях, агрегированных за последние три дня перед пожаром. Эффективность моделей измерялась на основе метрики точности f1-score на тестовой выборке, которая составляет 20% от исходного набора данных.

На первом этапе модель обучалась без грозовых признаков. Наибольшее значение метрики f1 на тестовой выборке было получено при использовании погодных признаков, агрегированных за последние 3 дня, что составило 0,93. При увеличении периода агрегации до 7 и 14 дней наблюдалось последовательное снижение точности, на тестовой выборке значение уменьшилось до 0,92.

На втором этапе модель была обучена с учетом грозовых явлений и проведено сравнение моделей с целью понимания, как наличие факторов о грозах влияет на точность модели. Максимальное значение f1 вновь было получено при 3-дневной агрегации и составило 0,94 на тестовой выборке для модели Random Forest. Однако с увеличением временного интервала до 7 и 14 дней метрика точности снижалась аналогичным образом, как и при обучении без грозовых признаков: до 0,938 при учете интервала 7 дней и 0,935 – 14 дней. Остальные модели показали менее точный результат: LightGBM – 0,938, ExtraTrees – 0,912, CatBoost – 0,931.

Необходимо отметить, что при добавлении грозовых признаков качество модели улучшилось на 0,01 по метрике f1 на тестовой выборке. Это говорит о том, что грозовые признаки имеют положительное, но не решающее влияние на точность прогнозирования.

По важности наибольшее влияние оказали признаки: засушливые дни, температура воздуха за последние 3-7 дней до пожара, влажность воздуха и почвы (3 и 7 дней) и наличие сухой грозы имеет значимую, но не ведущую роль. Низкая значимость бинарного признака грозы объясняется тем, что модель получает аналогичную информацию из более детальных погодных признаков, таких как дни засухи и температура воздуха. Таким образом, сам факт сухой грозы менее информативен, чем условия, при которых она возникла. Анализ SHAP выявил, что зоны за ветром (низкая ветровая экспозиция) на крутых склонах создают локально более сухие условия, повышающие вероятность возникновения грозовых пожаров во всех рассмотренных периодах. Дашборд с результатами исследования влияния грозовых признаков на прогнозирование пожара представлен на рисунке (рис. 3).

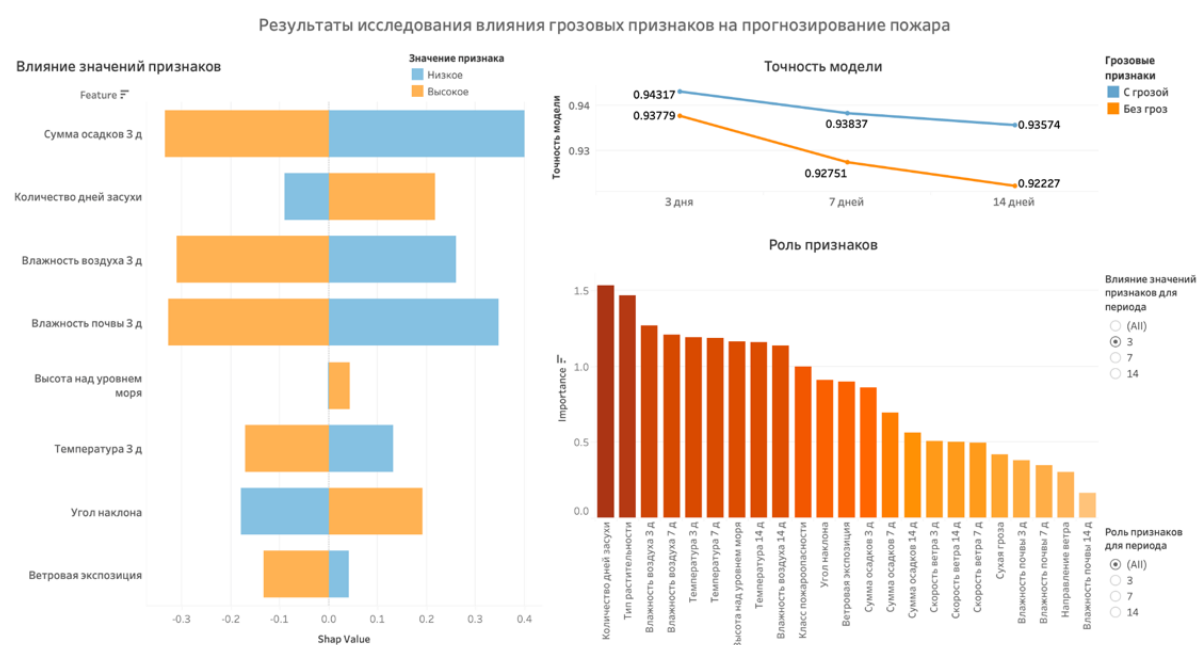


Рисунок 3: Визуализации важности признаков и точности прогнозирования риска пожаров

4. Заключение

Проведённое исследование подчёркивает сложность связи грозовой активности с возникновением лесных пожаров. Несмотря на интуитивную значимость молний как причины возгораний, наибольшую роль играют именно погодные условия, предшествующие грозе. Важным направлением дальнейших исследований может стать интеграция данных о длительности и мощности разряда молнии, а также их детального пространственного распределения. Отдельного внимания заслуживает исследование возможности возникновения вторичной грозовой активности над очагами крупных пожаров. Согласно исследованию [2], такие параметры могут значительно повысить качество прогнозирования.

Результаты исследования могут обеспечить разработку комплексных моделей, учитывающих динамическое взаимодействие между пожаром и атмосферными явлениями, и

существенно повысить эффективность прогнозирования и предупреждения лесных пожаров в условиях изменяющегося климата.

5. Благодарности

Работа выполнена в рамках гранта No 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды»).

6. Список литературы

- [1] Е.Г. Швецов, А.И. Сухинин, Е.И. Пономарев, Исследование влияния погодных условий на возникновение пожаров от гроз, Хвойные бореальной зоны том 25 1-2 (2008) 47–51.
- [2] Р.В. Васильев, М.А. Тащилин, А.В. Татарников, Сопоставление динамики термальных точек и зарегистрированных лесных пожаров с динамикой молниевых разрядов на Байкальской природной территории, Вычислительные технологии том 28 6 (2023) 37–45. doi:10.25743/ICT.2023.28.6.004.
- [3] А.Ю. Бибаева, Пирогенное преобразование геосистем Западного Прибайкалья: факторы, условия, современное состояние, Известия Иркутского государственного университета. Серия Науки о Земле 35 (2021) 3–18. doi: 10.26516/2073-3402.2021.35.3.
- [4] K. Virts, R. Holzworth, M. McCarthy, World Wide Lightning Location Network (WWLLN) Monthly Thunder Hour Data, Dataset available online from the NASA Global Hydrometeorology Resource Center DAAC, Huntsville, Alabama, U.S.A. 2024. doi: 10.5067/WWLLN/DATA101
- [5] Е.А. Лупян, Д.В. Лозин, С.А. Барталев, И.В. Балашов, Ф.В. Стыщенко, Оценка повреждений российских лесов пожарами в XXI веке на основе анализа интенсивности горения по данным прибора MODIS, Современные проблемы дистанционного зондирования Земли из космоса том 21 6 (2024) 233–249. doi: 10.21046/2070-7401-2024-21-6-233-249.
- [6] I.V. Bychkov, G.M. Ruzhnikov, R.K. Fedorov, A.K. Popova, Y.V. Avramenko, On classification of Sentinel-2 satellite images by a neural network ResNet-50, Computer Optics volume 47 3 (2023) 474–481.
- [7] S. Saha et. al. The NCEP Climate Forecast System Reanalysis, Bull. Amer. Meteor. Soc, volume 91 8 (2010) 1015–1057.
- [8] S. Saha et. al. The NCEP Climate Forecast System Version 2., J. Climate 27 (2014) 2185–2208.
- [9] L. Breiman, Random forests, Mach. Learn 45 (2001) 5–23. doi: 10.1023/A:1010933404324.