

Научная статья

УДК 004.8:614.8; DOI: 10.61260/2304-0130-2026-2-29-35

АНАЛИЗ СЛОЖНОСТИ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ ПРИРОДНЫХ ПОЖАРОВ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

✉ Гончаров Андрей Сергеевич.

Санкт-Петербургский университет ГПС МЧС России, Санкт-Петербург, Россия

✉ andreygoncharov04@mail.ru

Аннотация. В статье исследуется сложность задачи вероятностного прогнозирования возникновения природных пожаров методами машинного обучения. Проведен анализ обучающей выборки, включающей метеорологические, спутниковые и исторические данные, а также выполнена визуализация распределений и решена задача понижения размерности с помощью нелинейных методов t-SNE и UMAP. Продемонстрировано, что классы «пожар» и «нет пожара» в значительной степени перекрывают друг друга в признаковом пространстве. Полученные результаты говорят о необходимости применения сложных, нелинейных методов машинного обучения.

Ключевые слова: природные пожары, машинное обучение, t-SNE, UMAP, визуализация данных

Для цитирования: Гончаров А.С. Анализ сложности задачи прогнозирования природных пожаров методами машинного обучения // Природные и техногенные риски (физико-математические и прикладные аспекты). 2026. № 2 (58). С. 29-35. DOI: 10.61260/2304-0130-2026-2-29-35

Scientific article

COMPLEXITY ANALYSIS OF WILDFIRE FORECASTING TASK WITH MACHINE LEARNING

✉ Andrey S. Goncharov.

Saint-Petersburg university of State fire service of EMERCOM of Russia, Saint-Petersburg, Russia

✉ andreygoncharov04@mail.ru

Abstract. This article examines the complexity of probabilistic wildfire forecasting using machine learning. A training set including meteorological, satellite, and historical data is analyzed, distributions are visualized, and dimensionality reduction is performed using nonlinear t-SNE and UMAP methods. It is demonstrated that the "fire" and "no fire" classes significantly overlap in the feature space, suggesting the need for complex, nonlinear machine learning methods.

Keywords: wildfires, machine learning, t-SNE, UMAP, data visualization

For citation: Goncharov A.S. Complexity analysis of wildfire forecasting task with machine learning // Natural and Man-Made Risks (Physical, Mathematical, and Applied Aspects). 2026. No. 2 (58). pp. 29-35. DOI: 10.61260/2304-0130-2026-2-29-35.

Введение

Как известно, из многолетних наблюдений и различных исследований, лесные пожары возникают из-за влияния множества различных факторов, а именно: метеорологических условий, типа растительности, рельефа местности, антропогенной нагрузки, и, к сожалению, их практически невозможно описать линейными моделям [1–3]. Именно поэтому задачи вероятностного прогнозирования различных физических и природных процессов, в том числе и прогнозирования природных пожаров, считаются одними из наиболее сложных в прикладном машинном обучении [4–6]. Стандартные физические индексы пожарной опасности довольно ограничены и поэтому способны давать лишь приблизительную оценку, а методы машинного обучения позволяют выявлять глубокие нелинейные зависимости, но при условии хорошего качества и подходящей структуры данных [7–9].

Целью исследования является анализ структуры данных, их сложности, а также выборе алгоритмов машинного обучения наиболее подходящих для решения задачи прогнозирования природных пожаров. Для этого применены два широко известных метода визуализации данных с помощью понижения размерности – t-SNE и UMAP, они позволяют определить, насколько классы «пожар» и «нет пожара» разделимы в признаковом пространстве и какие алгоритмы обучения использовать [10, 11]. Такой подход позволяет как оценить сложность решаемой задачи, так и дать обоснованные рекомендации по выбору архитектуры будущей системы прогнозирования.

Методы исследования

Начать необходимо с выбора исходных данных, в данной работе были использованы данные из открытых источников, они включали в себя метеорологические данные с восемью метеорологических станций Сибири и Дальнего Востока, а также данные дистанционного зондирования земли с информационного ресурса NASA FIRMS MODIS за период с 2019 по 2024 гг. Была построена обучающая выборка для прогнозирования возникновения природных пожаров с упреждением в 7 суток. Все переменные были сведены в единое признаковое пространство из 140 характеристик, включающее экспоненциально взвешенные средние, лаговые значения метеопараметров и спутниковых данных, сезонные компоненты. Целевая переменная определялась как индикатор возникновения нового пожара в ближайшие 7 суток.

Итоговый объем выборки составил 9896 наблюдений, с долей положительного класса 44%.

Для анализа данных и их визуализации с проецированием на плоскость использовали два алгоритма нелинейного понижения размерности t-SNE и UMAP. Метод t-SNE сжимает гиперпространство признаков, на основе вероятностной близости объектов и наиболее эффективно сохраняет локальные соседства и зависимости, что позволяет выявлять кластеры и области смешения, перекрытия классов [10]. Метод UMAP основан на топологическом анализе данных, но в упрощенном графовом представлении данных и стремится сохранить локальную и глобальную структуру выборки, создавая компактные и понятные кластеры [11].

Перед построением результатов анализа на графиках, было необходимо поставить оба метода в одинаковые условия, то есть стандартизировать. Для этого в настройках метода t-SNE был выбран параметр perplexity равный 30, этот параметр можно представить как количество ближайших соседей, которые алгоритм будет пытаться сохранить при понижении размерности пространства к двумерному, для наилучшей визуализации. Значение параметра 30 является стандартным и хорошо работает для выборок в несколько тысяч наблюдений [12]. Для UMAP использовались основные настройки по умолчанию n_neighbors равный 15 и min_dist равный 0.1. Первый параметр в чистом виде количество ближайших соседей, второй определяет, насколько плотно точки могут располагаться друг к другу [13].

Оба метода запускались на случайной подвыборке из 5000 наблюдений, для более быстрого вычисления и получения результатов, основные закономерности при этом не утрачиваются. Полный список параметров представлен на рис. 1.

```
# Размер подвыборки
N_SAMPLES = 5000
# Параметры методов
tsne_params = {
    'n_components': 2,
    'random_state': 42,
    'perplexity': 30,
    'init': 'pca',          # по умолчанию в sklearn
    'learning_rate': 'auto',
    'max_iter': 1000       # значение по умолчанию
}

umap_params = {
    'n_components': 2,
    'random_state': 42,
    'n_neighbors': 15,     # по умолчанию в umap-learn
    'min_dist': 0.1,      # по умолчанию
    'metric': 'euclidean',
    'n_jobs': 1
}
```

Рис. 1. Листинг программы

Результаты

После завершения работы алгоритмов были получены следующие результаты, представленные на рис. 2 и 3.

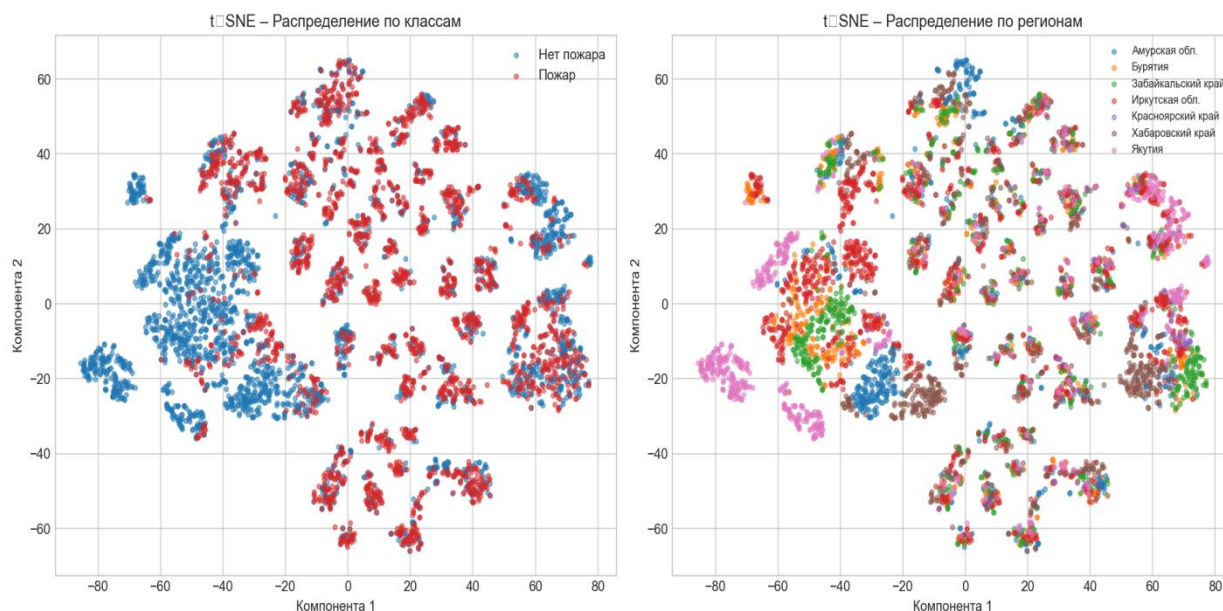


Рис. 2. Результаты визуализации данных методом t-SNE

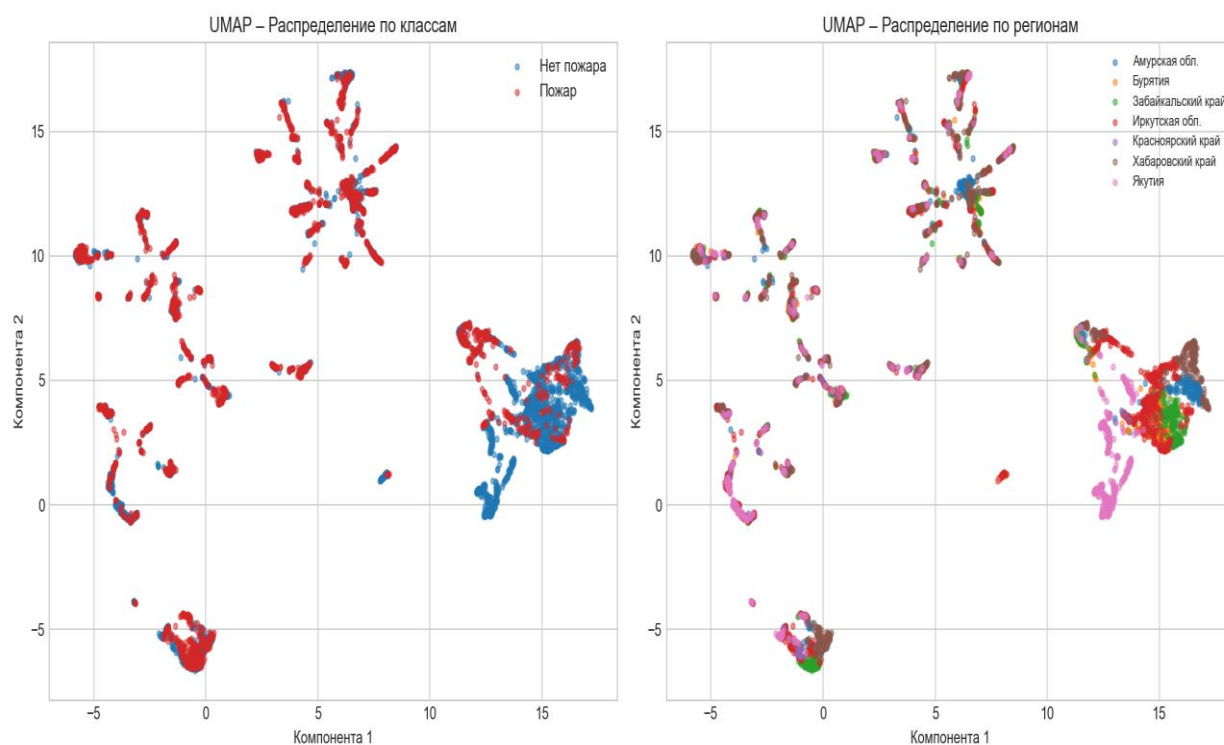


Рис. 3. Результаты визуализации данных методом UMAP

Как видно из рис. 2 и 3, явно выделяется сильное перемешивание синих (нет пожара) и красных (пожар) точек, о явной границе между классами говорить не приходится, а это означает лишь то, что никакая даже самая совершенная модель не сможет достичь идеальной точности, так как в данных присутствуют объекты разных классов при почти одинаковых признаках, исходя из этого, вывод напрашивается сам собой – задача объективно сложная и линейные модели здесь не покажут высокой эффективности.

Говоря о региональной структуре (правые графики рис. 2 и 3), можно сделать вывод, что регионы образуют отчетливые, хотя и немного перекрывающиеся, кластеры. Особенно выделяется Якутия, из-за особенностей климата и насаждений, точки формируют почти обособленный кластер. Также видны близко расположенные друг к другу кластеры разных регионов, что говорит о существовании общего климатического характера, а также похожий тип насаждений, вследствие чего и близкие причины возникновения природных пожаров.

Также было проведено сравнение методов, по его результатам UMAP оказался более предпочтительным, так как дает более компактную и легко читаемую картину, кластеры более аккуратные, отчетливее просматривается глобальная структура, t-SNE немного сложнее интерпретировать, так как он разбивает локальные группы, в целом это объясняется различиями в основе построения этих методов, но в целом можно сказать, что метод UMAP более удобный инструмент для разведочного анализа сложных данных.

Полученные в ходе исследования результаты имеют практическую значимость, так как становится очевидно, что классы линейно неразделимы и поэтому приоритетом при построении прогностических моделей должны быть полнота правильно предсказанных данных, способность моделей корректно ранжировать дни по степени риска и выдавать хорошо откалиброванные вероятности возникновения природных пожаров. Наличие региональных кластеров и их частичное пересечение говорит о том, что модели обязаны учитывать географические факторы. В таких условиях неприменимы методы типа регрессии, логистической регрессии и методы опорных векторов, здесь хорошо показывают себя ансамблевые методы и градиентный бустинг [14], а также рекуррентные нейронные сети, в частности LSTM, которые способны учитывать долгосрочные зависимости без явного конструирования лаговых признаков [15].

Заключение

Проведенный визуальный анализ гиперпространства признаков задачи прогнозирования природных пожаров показал, что классы имеют высокую степень перекрытия, при одинаковых условиях существуют оба класса, регионы формируют выраженные, но все же пересекающиеся группы, все это объясняет почему простые линейные модели неэффективны, и обосновывает применение более сложных архитектур типа градиентного бустинга и рекуррентных нейронных сетей [14, 15]. Проведенное сравнение методов снижения размерности и визуализации выявило незначительное преимущество UMAP для задач нашего типа с точки зрения наглядности и компактности представления данных. Полученные результаты на практике определяют выбор архитектур прогностических моделей.

Список источников

1. Матиев Р.Т., Матвеев А.В., Таранцев А.А. Модель поддержки принятия решений при реагировании на лесные пожары в горной местности с использованием нечеткой логики // Инженерный вестник Дона. 2025. № 12 (132). EDN LCTAWP.
2. Медведев Д.В., Матвеев А.В. Алгоритмы интеллектуальной поддержки принятия управленческих решений при угрозах лесных пожаров // Научно-аналитический журнал «Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России». 2025. № 2. С. 35-48. DOI: 10.61260/2218-130X-2025-2-35-48. EDN OKGHLE.
3. Матвеев А.В., Матиев Р.Т. Принятие решений при пожарах в горной местности: Сравнительный анализ методов мониторинга // Национальная безопасность и стратегическое планирование. 2023. № 2(42). С. 76-90. DOI: 10.37468/2307-1400-2023-2-76-90. EDN QPRUWC.
4. A review of machine learning applications in wildfire science and management / P. Jain [et al] // Environmental Reviews. 2020. V. 28. No 4. P. 478-505. DOI: 10.1139/er-2020-0019.
5. Wildfire Risk Prediction: A Survey of Recent Advances Using Deep Learning Techniques / Z Xu [et al.] // 2024. arXiv:2405.01607v4
6. Медведев Д.В., Матвеев А.В. Информационная модель поддержки принятия решений по реагированию на ландшафтные пожары // «Сибирский пожарно-спасательный вестник». 2025. № 1(36). С. 117-125. DOI: 10.34987/vestnik.sibpsa.2025.22.36.011. EDN PKSBTR.
7. Медведев Д.В., Матвеев А.В., Смирнов А.С. Применение модели логистической регрессии при принятии решений по определению количества привлекаемых сил на ликвидацию лесных пожаров // Пожаровзрывобезопасность. 2024. Т. 33. № 4. С. 84-96. DOI: 10.22227/0869-7493.2024.33.04.84-96. EDN MJLVTY.
8. Медведев Д.В., Матвеев А.В., Дмитриева А.И. Применение нейро-нечеткой системы ANFIS при прогнозировании площади лесных пожаров // Пожарная безопасность: современные вызовы. Проблемы и пути решения: сб. материалов Всерос. науч.-практ. конф. СПб.: С.-Петерб. ун-т ГПС МЧС России, 2024. С. 43-48. EDN PFZVXO.
9. Медведев Д. В., Матвеев А. В. Свидетельство о государственной регистрации программы для ЭВМ № 2024682958 Российская Федерация. Нейро-нечеткая система ANFIS для прогнозирования показателей лесных пожаров: № 202468121: заявл. 13.09.2024: опубл. 01.10.2024 // С.-Петерб. ун-т ГПС МЧС России. EDN MMAXOR.
10. Van der Maaten L., Hinton G. Visualizing data using t-SNE // Journal of machine learning research. 2008. V. 9. No. 11.
11. McInnes L., Healy J., Melville J. Umap: Uniform manifold approximation and projection for dimension reduction // arXiv preprint arXiv: 1802.03426. 2018. DOI: 10.48550/arXiv.1802.03426
12. Scikit-learn: T SNE. URL: <https://scikit-learn.org/stable/modules/manifold.html#t-distributed-stochastic-neighbor-embedding-t-sne> (дата обращения: 06.02.2026).
13. Umap-learn documentation. URL: <https://umap-learn.readthedocs.io/> (дата обращения: 06.02.2026).

14. CatBoost: unbiased boosting with categorical features / L. Prokhorenkova [et al.] // *Advances in Neural Information Processing Systems*. 2018. Vol. 31.
15. Hochreiter S., Schmidhuber J. Long Short Term Memory // *Neural Computation*. 1997. Vol. 9. № 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

References

1. Matiev R.T., Matveev A.V., Tarancev A.A. Model` podderzhki prinyatiya reshenij pri reagirovanii na lesnye pozhary v gornoj mestnosti s ispol`zovaniem nechetkoj logiki // *Inzhenernyj vestnik Dona*. 2025. № 12(132). EDN LCTAWP.
2. Medvedev D.V., Matveev A.V. Algoritmy intellektual`noj podderzhki prinyatiya upravlencheskix reshenij pri ugrozax lesny`x pozharov // *Nauchno-analiticheskij zhurnal Vestnik Sankt-Peterburgskogo universiteta Gosudarstvennoj protivopozharnoj sluzhby` MChS Rossii*. 2025. № 2. S. 35-48. DOI: 10.61260/2218-130X-2025-2-35-48. EDN OKGHLE.
3. Matveev A.V., Matiev R.T. Prinyatie reshenij pri pozharax v gornoj mestnosti: Sravnitel`nyj analiz metodov monitoringa // *Nacional`naya bezopasnost` i strategicheskoe planirovanie*. 2023. № 2(42). S. 76-90. DOI: 10.37468/2307-1400-2023-2-76-90. EDN QPRUWC.
4. P. Jain [et al.] A review of machine learning applications in wildfire science and management // *Environmental Reviews*. 2020. V. 28. No 4. P. 478-505. DOI: 10.1139/er-2020-0019
5. Wildfire Risk Prediction: A Survey of Recent Advances Using Deep Learning Techniques / Z Xu [et al.] // 2024. arXiv:2405.01607v4.
6. Medvedev D.V., Matveev A.V. Informacionnaya model` podderzhki prinyatiya reshenij po reagirovaniyu na landshaftnye pozhary // *Sibirskij pozharo-spasatel`nyj vestnik*. 2025. № 1(36). S. 117-125. DOI: 10.34987/vestnik.sibpsa.2025.22.36.011. EDN PKSBTR.
7. Medvedev D.V., Matveev A.V., Smirnov A.S. Primenenie modeli logisticheskoy regressii pri prinyatii reshenij po opredeleniyu kolichestva privlekaemyx sil na likvidaciyu lesnyx pozharov // *Pozharovzryvbezopasnost`*. 2024. T. 33. № 4. S. 84-96. DOI: 10.22227/0869-7493.2024.33.04.84-96. EDN MJLVTY.
8. Medvedev D.V., Matveev A.V., Dmitrieva A.I. Primenenie nejro-nechetkoj sistemy ANFIS pri prognozirovanii ploshhadi lesny`x pozharov // *Pozharnaya bezopasnost`: sovremennye vyzovy. Problemy i puti resheniya: sb. materialy` Vseros.j nauch.-prakt. konf., Sankt-Peterburg, 18 aprelya 2024 goda. Sankt-Peterburg: Sankt-Peterburgskij universitet GPS MChS Rossii, 2024. S. 43-48. EDN PFZVXO.*
9. Svidetel'stvo o gosudarstvennoj registracii programmy` dlya E`VM № 2024682958 Rossijskaya Federaciya. Nejro-nechetkaya sistema ANFIS dlya prognozirovaniya pokazatelej lesny`x pozharov : № 2024681214 : zayavl. 13.09.2024 : opubl. 01.10.2024 / D.V. Medvedev, A.V. Matveev; zayavitel` Sankt-Peterburgskij universitet GPS MChS Rossii. EDN MMAXOR.
10. Van der Maaten L., Hinton G. Visualizing data using t-SNE // *Journal of machine learning research*. 2008. V. 9. No. 11.
11. McInnes L., Healy J., Melville J. Umap: Uniform manifold approximation and projection for dimension reduction // *arXiv preprint arXiv: 1802.03426*. 2018. DOI 10.48550/arXiv.1802.03426
12. Scikit-learn: T SNE. URL: <https://scikit-learn.org/stable/modules/manifold.html#t-distributed-stochastic-neighbor-embedding-t-sne/> (data obrashcheniya: 06.02.2026).
13. Umap-learn documentation. URL: <https://umap-learn.readthedocs.io/> (data obrashcheniya: 06.02.2026).
14. CatBoost: unbiased boosting with categorical features / L. Prokhorenkova [et al.] // *Advances in Neural Information Processing Systems*. 2018. Vol. 31.
15. Hochreiter S., Schmidhuber J. Long Short Term Memory // *Neural Computation*. 1997. Vol. 9. № 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735

Информация о статье:

Статья поступила в редакцию: 24.02.2026; одобрена после рецензирования: 26.05.2026;
принята к публикации: 29.05.2026

Information about the article:

The article was submitted to the editorial office: 24.02.2026; approved after review: 26.05.2026;
accepted for publication: 29.05.2026

Информация об авторах:

Гончаров Андрей Сергеевич, кафедра системного анализа и антикризисного управления Санкт-Петербургского университета ГПС МЧС России (196105, Санкт-Петербург, Московский пр., д. 149),
e-mail: andreygoncharov04@mail.ru

Information about the authors:

Goncharov Andrey S. department of systems analysis and crisis management at the Saint-Petersburg university of State fire service of EMERCOM of Russia (196105, Saint-Petersburg, Moskovsky ave., 149),
e-mail: andreygoncharov04@mail.ru