

Мариупольский Тимофей Игоревич

студент

Панфилов Илья Александрович

канд. техн. наук, доцент

ФГБОУ ВО «Сибирский государственный университет науки
и технологий им. академика М.Ф. Решетнева»

г. Красноярск, Красноярский край

МОДЕЛИРОВАНИЕ ВЕЛИЧИНЫ УЩЕРБА ПО ДАННЫМ ПРИРОДНЫХ И АНТРОПОГЕННЫХ КАТАСТРОФ

Аннотация: в статье рассмотрено решение задачи выявления неочевидных связей и закономерностей между факторами пожаров. Проведена категоризация пожаров по наличию пострадавших в пожарах, а также построение регрессионных моделей для расчета целевого фактора – прямого ущерба.

Ключевые слова: анализ данных, машинное обучение, природные пожары, природные катастрофы, антропогенные катастрофы.

Введение.

Природные и антропогенные катастрофы представляют собой серьезную угрозу для безопасности людей, а также наносят значительный экономический ущерб. В последние годы участились случаи таких катастроф, что подчеркивает необходимость их детального анализа и разработки эффективных стратегий управления рисками [1]. Данная проблема особенно важна в контексте происхождения подобных катастроф на территории субъектов Российской Федерации, которые обладают разнообразными климатическими и географическими особенностями.

Цель исследования.

Основная цель данной работы – разработать методологию для анализа данных о катастрофах, чтобы выявить ключевые факторы и паттерны, влияющие на последствия этих событий, построить модель для проведения классификации пожаров по наличию пострадавших, а также построение регрессионных моделей с

целевым фактором прямого ущерба [2]. Для этого были поставлены несколько задач: сбор данных, их предобработка, анализ и построение моделей.

Сбор и предобработка данных.

Сотрудниками ВНИИ Пожарной охраны были собраны данные о природных и антропогенных пожарах за 10 лет, которые были использованы в данной работе. Всего была представлена информация о более чем 3 млн. пожаров. Была проведена предобработка данных, а именно заполнение пропусков, удаление аномалий и нормализация данных для проведения дальнейшего анализа [3]. В таблице 1 представлены факторы из набора данных, по которым будет осуществляться анализ данных.

Таблица 1

Расшифровка отобранных для проведения анализа факторов

Фактор	Расшифровка фактора
F6	Вид населённого пункта
F7	Вид пожарной охраны населённого пункта
F8	Организационно-правовая форма
F10	Ведомственная принадлежность
F11	Тип предприятия, организации, учреждения
F12	Объект пожара
F14	Этажность здания
F15	Этаж, на котором возник пожар
F16	Степень огнестойкости
F17	Место возникновения
F17a	Тип места
F19	Причина пожара
F26	Расстояние до пожарной части, км
F27	Погибло людей: всего
F30	Получили травмы: всего
F36	Прямой ущерб
F39	Уничтожено: Строений
F41	Жилых квартир
F44	Кв.м. площади
F45	Автотракторной техники, ед.
F46	Другой техники, ед.
F56	Спасено: людей
F60	Материальных ценностей, руб
F72	Условия, способствовавшие развитию пожара
F75	Участники тушения пожара
F78	Техника
F83	Количество техники, ед
F92	Количество пожарных стволов, ед
F94	Огнетушащие средства

F100	Водоисточники
F106	Результаты работы

Инструменты и технологии для анализа данных.

Для классификации данных и регрессионного анализа были использованы модели из scikit-learn, предоставляющие возможность обучения моделей на данных и их последующего применения для прогнозирования. Для проведения регрессионного анализа были использованы соответствующие инструменты из библиотеки scikit-learn, такие как LinearRegressor, Ridge, Lasso, ElasticNet, RandomForestRegressor, GradientBoostingRegressor. Эти методы позволяют построить регрессионные модели, а mean_squared_error и r2_score позволяют оценить качество построенных регрессионных моделей. Для визуализации данных были использованы библиотеки matplotlib, plotly, seaborn позволяющие создавать интерактивные и информативные визуализации, улучшая понимание данных и результатов анализа [4].

Применение инструментов визуализации и машинного обучения.

Использованы методы машинного обучения для анализа данных о пожарах. В частности, разработана модель для классификации пожаров с пострадавшими и без пострадавших, а также были построены регрессионные модели, чтобы понять, какие факторы влияют на ущерб.

Была разработана модель для классификации пожаров с пострадавшими и без пострадавших, а также были построены регрессионные модели, чтобы понять, какие факторы влияют на ущерб.

По результату работы алгоритма классификации, были получены следующие результаты (рисунок 1):

```

Classification Report:
              precision    recall  f1-score   support

     0           1.00      0.78      0.87     301864
     1           0.12      0.92      0.21       9902

 accuracy          0.78     311766
 macro avg          0.56     311766
 weighted avg       0.97     311766

Confusion Matrix:
[[235222  66642]
 [   823  9079]]

```

Рис. 1. Результаты классификации

Точность работы алгоритма составила 78% (235222 случаев – верно, 66642 – неверно). Итоги работы алгоритма классификации признаны удовлетворительными.

На рисунке 1 представлена тепловая карта по всем факторам пожаров. В результате были отобраны наиболее значимые факторы для построения регрессионных моделей.

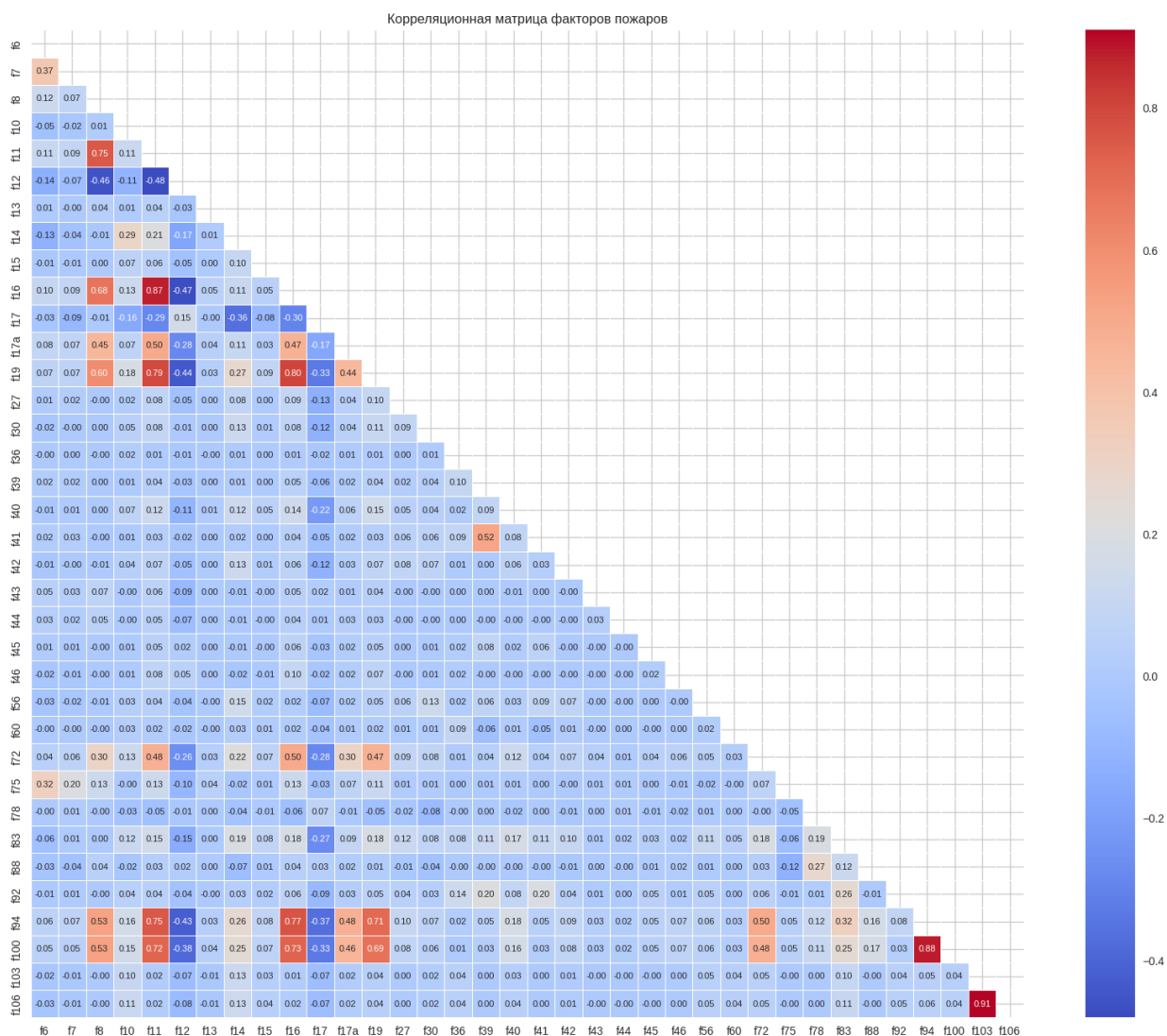


Рис. 2. Корреляционная матрица по всему датасету

По итогам работы моделей регрессии, были получены следующие результаты (таблица 2):

Таблица 2

Результаты работы регрессионных моделей

<i>Model</i>	<i>Linear Regression</i>	<i>Ridge Regression</i>	<i>Lasso Regression</i>	<i>ElasticNet</i>	<i>Decision Tree</i>	<i>Random Forest</i>	<i>Gradient Boosting</i>
<i>MS</i>	24205731	24249173	24205768	25202941	25800566	24907612	26176191
<i>E</i>	0578.929	9215.846	9841.529	3669.412	6738.606	7955.113	3935.150
<i>R²</i>	0.053516	0.051818	0.053515	0.014524	-0.00884	0.026072	-0.02352

Как видно по итогам работы моделей, ни одна из них не представляет удовлетворительных результатов. Обоснованность влияния данных на фактор прямого ущерба достигает максимум 5%, что совершенно не удовлетворяет целям.

Заключение.

Проведенный анализ подтвердил практическую значимость и эффективность использования современных методов анализа данных для исследования природных и антропогенных катастроф. Результаты демонстрируют, что комплексный подход к сбору, обработке и анализу данных может существенно повысить качество принятия решений и разработку стратегий по управлению рисками катастроф.

В дальнейшем исследовании планируется использовать более сложные методы машинного обучения. Это должно повысить точность прогнозов и способствовать разработке эффективных стратегий управления рисками катастроф.

Список литературы

1. Савонин С.В. Пожары и их последствия. анализ статистических данных / С.В. Савонин, Т.В. Украинцева, А.С. Мазур [и др.] // Известия Санкт-Петербургского государственного технологического института (технического университета). – 2019. – №50 (76). – С. 110–114. – EDN TNBVHN
2. Мариупольский Т.И. Анализ данных антропогенных пожаров / Т.И. Мариупольский, В.В. Ничепорчук, Д.В. Кустов [и др.] // Научно-технический вестник Поволжья. – 2024. – №6. – С. 80–82. – EDN DETKCI
3. Лахвицкий Г.Н. Анализ влияния профилактической работы на динамику пожаров на основе статистических данных / Г.Н. Лахвицкий // Сибирский пожарно-спасательный вестник. – 2023. – №3 (30). – С. 65–72. – DOI 10.34987/vestnik.sibpsa.2023.30.3.007. – EDN JYHOSB

4. Панфилов И.А. Разработка системы прогнозирования сроков навигации на реках Енисейского бассейна / И.А. Панфилов, Е.И. Сивцова, С.Е. Маегов [и др.] // Перспективы науки. – 2022. – №8 (155). – С. 26–30. – EDN GBSNFC