

Научная статья
УДК 614.842
doi: 10.34987/vestnik.sibpsa.2023.19.99.005

Применение технологий машинного обучения для исследования характеристик пожаров

Константин Сергеевич Власов¹
Александр Алексеевич Порошин²
Ольга Сергеевна Маторина³

ВНИИПО МЧС России, Балашиха, Россия

¹<https://orcid.org/0000-0003-2499-169X>

²<https://orcid.org/0000-0001-9849-7024>

³<https://orcid.org/0000-0003-3962-6492>

Автор ответственный за переписку: *Ольга Сергеевна Маторина, odp1313@yandex.ru*

Аннотация. В статье рассмотрены вопросы применения технологий машинного обучения (ML) для исследования характеристик пожаров. Рассмотрены основные теоретические и практические положения технологии ML применительно к анализу параметров тушения пожаров, с учетом таких их видов как «ординарный» и «затяжной пожар». На примере пожаров, происшедших в Магаданской, Московской областях и Забайкальском крае, показана процедура определения факторов и их взаимосвязей, оказывающие значительное влияние на процесс пожаротушения.

Ключевые слова: машинное обучение, ординарный и затяжной пожар, статистическая выборка, функция правдоподобия

Для цитирования: Власов К.С., Порошин А.А., Маторина О.С. Применение технологий машинного обучения для исследования характеристик пожаров // Сибирский пожарно-спасательный вестник. 2023. № 2(29). С. 80-87.
<https://doi.org/10.34987/vestnik.sibpsa.2023.19.99.005>

APPLICATION OF MACHINE LEARNING TECHNOLOGIES TO STUDY THE CHARACTERISTICS OF FIRES

Konstantin S. Vlasov¹
Aleksandr A. Poroshin²
Olga S. Matorina³

¹<https://orcid.org/0000-0003-2499-169X>

²<https://orcid.org/0000-0001-9849-7024>

³<https://orcid.org/0000-0003-3962-6492>

VNIPO EMERCOM of Russia, Balashikha, Russia

Corresponding author: *Olga S. Matorina, odp1313@yandex.ru*

Abstract. The article discusses the application of machine learning (ML) technologies to study the characteristics of fires. The main theoretical and practical provisions of ML technology are considered in relation to the analysis of fire extinguishing parameters, taking into account their types such as “ordinary” and “protracted fire”. Using the example of fires that occurred in the Magadan,

Moscow regions and the Trans-Baikal Territory, the procedure for determining factors and their interrelations that have a significant impact on the fire extinguishing process is shown.

Keywords: machine learning, ordinary and protracted fire, statistical sampling, likelihood function

For citation: Vlasov K.S., Poroshin A.A., Matorina O.S., Application of machine learning technologies to study the characteristics of fires// Siberian Fire and Rescue Bulletin.2023;2(29): 80-87. (In Russ.). [https:// doi.org/10.34987/vestnik.sibpsa.2023.19.99.005](https://doi.org/10.34987/vestnik.sibpsa.2023.19.99.005).

Одной из форм искусственного интеллекта [1] является машинное обучение (Machine Learning – далее ML), которое позволяет выявлять определенные закономерности в данных, описывающих изучаемую предметную область. На основе выявленных закономерностей создаются математические модели для прогнозирования и построения аналитических сценариев развития исследуемых явлений. Технологии ML нашли широкое применение в различных отраслях научных знаний и активно используются на практике [2,3]. В этой связи целесообразно рассмотреть вопрос о применении технологий ML в области пожарной безопасности, в частности для исследования характеристик пожаров. Наибольший интерес в таком исследовании представляют тактически сложные пожары, для ликвидации которых привлекаются значительные силы и средства пожарно-спасательных гарнизонов (далее – ПСГ).

Оценка возможностей сил и средств территориальных или местных ПСГ по ликвидации тактически сложных пожаров определяет объективность сложившейся практики нормирования в области ресурсного обеспечения пожарно-спасательных подразделений (далее – ПСП) с учетом технико-экономических характеристик объектов защиты и природно-климатических особенностей в соответствующем субъекте Российской Федерации.

Учет особенностей субъектов Российской Федерации хорошо иллюстрирует следующий пример. На Рис приведены статистические распределения по пожарам за 2021 год в Магаданской, Московской областях и Забайкальском крае, с учетом таких показателей тушения как: время занятости ПСП и количество мобильной пожарной техники, привлекаемой для ликвидации пожаров. Распределения приведены на фоне ϵ – Гауссовского шума. Как видно из рисунка 1, в зависимости от особенностей каждого из рассматриваемых субъектов Российской Федерации, рассматриваемые параметры ликвидации пожаров существенно отличаются друг от друга. Однако, в целом анализ показывает, что значительное число пожаров в рассматриваемых регионах ликвидируются за время, не превышающее одного часа. На рисунке 2 приведен пример распределения показателя времени занятости на пожаре ПСП в Московской области.

Пожары, для ликвидации которых потребовалось менее одного часа, в дальнейшем будем называть «ординарными».

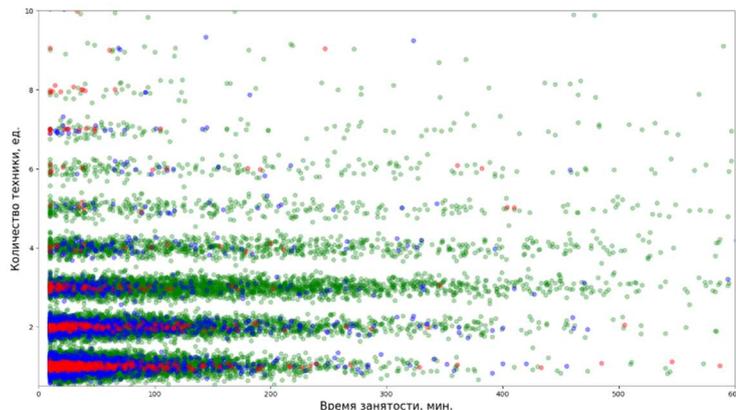


Рис.1. Распределение числа пожаров, произошедших в 2021 году в Магаданской (красные точки), Московской (зеленые точки) и Забайкальском крае (синие точки) по времени занятости ПСП (мин.) и количеству мобильной пожарной техники (ед.)

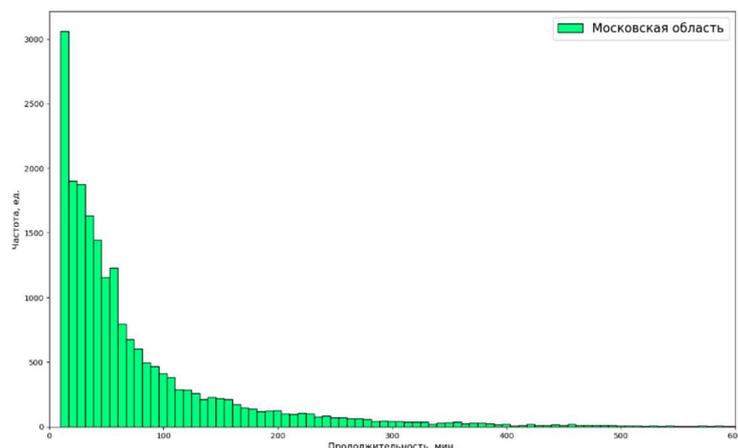


Рис.2. Частота распределения пожаров, произошедших в 2021 году в Московской по времени занятости (мин.)

В случае превышения времени ликвидации пожара более одного часа, можно предположить, что данный пожар представлял повышенную тактическую сложность для ПСП для его ликвидации. Вместе с тем, длительность тушения пожаров не всегда является достоверным фактом, подтверждающим наличие тактической сложности по тушению для ПСП. Например, горение (тление) торфа или тушение горючих материалов, скопившихся в труднодоступных для подачи огнетушащих веществ местах, может длиться сутками. Как показывает практика, при этом не возникает необходимости привлечения большого количества сил и средств ПСГ.

В общем виде, определено [4], что в ситуациях, когда для успешной ликвидации горения требуется высокая интенсивность действий участников тушения пожара с использованием мобильной специальной техники в условиях воздействия опасных факторов пожара в течение длительного времени, то такие пожары относятся к затяжным. Время длительности затяжных пожаров принято фиксировать от двух часов и более.

Решение задачи прогнозирования возможности перехода ординарного пожара в стадию затяжного, с определением его характеристических признаков, представляет определенный интерес с позиций выявления особенностей параметров их ликвидации в различных условиях субъектов Российской Федерации.

Решения данной задачи было рассмотрено на основе применения технологии ML. Была сформирована статистическая выборка $\{N\}$ по трем вышеназванным регионам Российской Федерации, состоящая из 26389 наблюдений (образцов)¹. Определены 17 показателей (признаков) по оперативной деятельности ПСП. Показатели отобраны с учетом признаков прогнозирования вариантов развития пожара. Описание процесса подготовки данных, разработки модели, проектирования признаков и написания программного кода на языке Python приводится в файле Jupyter Notebook расположенного в репозитории [5].

Применен алгоритм ML «обучение с учителем» [0], который предполагает организацию набора исследуемых данных в виде коллекции размеченных образцов вида

$$\left\{ \left(x_i^{(j)}, y_i \right) \right\}_{i=1}^N \quad (1)$$

Каждый элемент x_i из статистическая выборка $\{N\}$ представляет собой вектор признаков отдельного пожара (образца), в котором $j = 1, \dots, N$ содержатся характеристики единичного признака. Например, $x^{(0)}$ – код региона Российской Федерации, $x^{(1)}$ – тип населенного пункта, где произошел пожар и т.д. Метка y_i является элементом конечного множества классов $\{1, 2, \dots, C\}$. В нашем случае это признак ординарного или затяжного пожара $\{0,1\}$. Цель выполнения

¹ В технологиях ML для наблюдений используется термин «образец», а для показателей – «признак».

применяемого алгоритма: на основе вектора признаков $x_i^{(j)}$ определить соответствующую метку y .

Для выбора ведущих факторов из статистической выборки $\{N\}$, на основе корреляционной матрицы (см. ее фрагмент на Рис.3) определены наличие и сила связей между каждой парой исследуемых признаков. Так, признаки F16 (Степень огнестойкости) и F12 (Объект пожара) имеют сильную отрицательную корреляционную связь. Данная ситуация объясняется сходством признаков. То есть степень огнестойкости зависит от вида объекта защиты. Это означает, что для прогнозирования характеристик затяжного пожара один из показателей является избыточным, и может быть удален из выборки $\{N\}$. Уточненная корреляционная матрица взаимосвязей признаков тушения пожара приведена на рис.4.

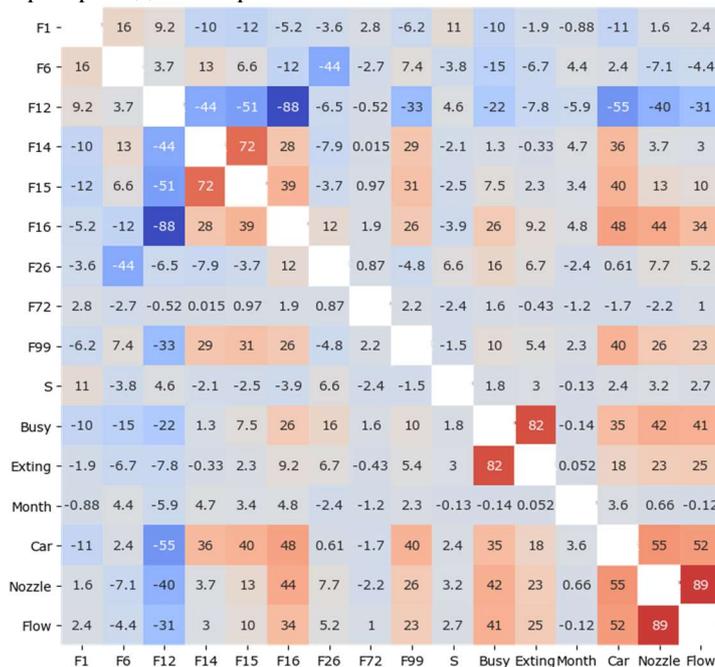


Рис.3. Корреляционная матрица признаков пожара: F1 - субъект РФ; F6 - вид населенного пункта; F12 - объект пожара; F14 - этажность здания; F15 - этаж, на котором возник пожар; F16 - степень огнестойкости; F26 - расстояние до пожарной части; F72 - условия, способствующие развитию пожара; F99 - использование СИЗОД; S - площадь пожара; Busy - время занятости на пожаре; Exting - время подачи стволов; Month - номер месяца года; Car - общее количество техники, ед.; Nozzle - общее количество стволов, ед.; Flow – расчетный расход воды на тушение, л/с

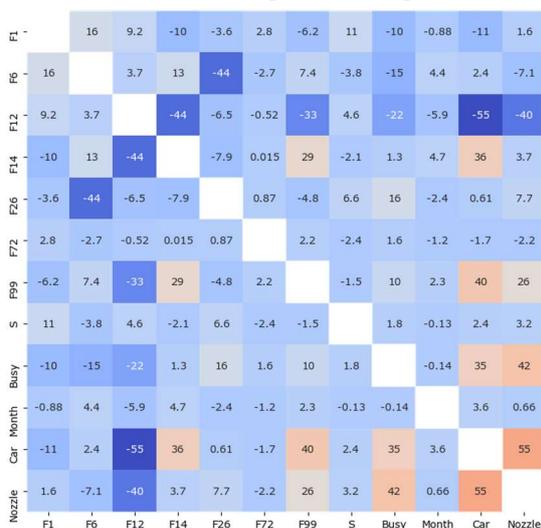


Рис.4. Уточненная корреляционная матрица признаков тушения пожара

Исследование выборки данных на наличие их выбросов (даже без признаков типа «Busy» и «S», явно имеющих очень большие разбросы) показывает, что признаки так же неоднородны (см. Рис.5). Для устранения неоднородности признаков проводится процедура их нормализации, условием которой, является обеспечение для всех признаков равных возможностей влияния на результирующий фактор. В работах [0, 6] отмечено, что в этом случае алгоритмы ML работают лучше и быстрее сходятся. Не существует единого универсального метода нормализации. В основном применяются такие методы как: «Минимакс» (MinMaxScaler), «Z-масштабирование» (StandardScaler) или «Робастная нормализация» (RobustScaler). На основании сравнительного анализа модифицированных распределений признаков (см. Рис.6) и рекомендаций, приведенных в работе [6], в настоящих исследованиях применен метод «Z-масштабирование» (StandardScaler). Итоговое распределение числовых признаков тушения пожаров после проведения процедуры нормализации приведено на рис.7.

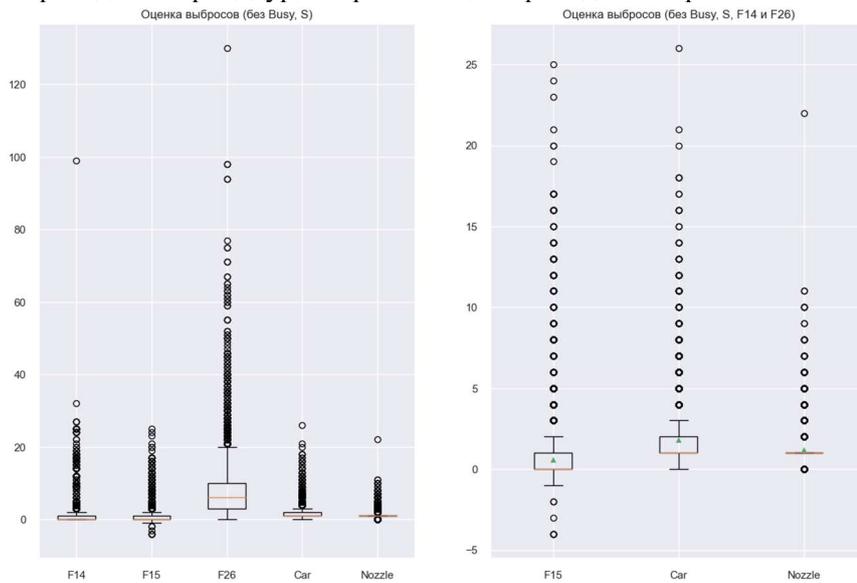
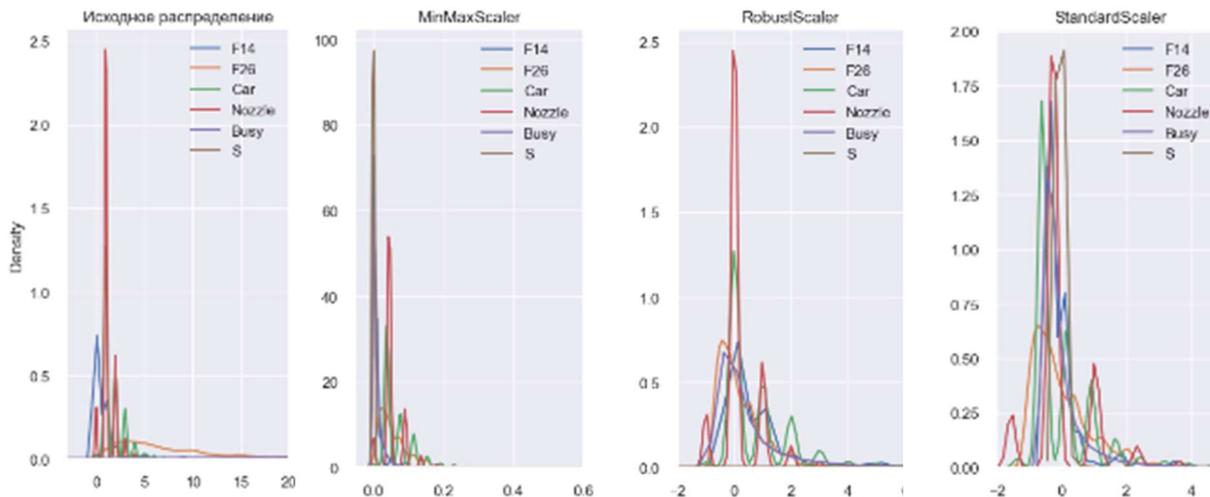


Рис.5. Определение выбросов исследуемых признаков. Оранжевая полоска – медиана признака. Вершина треугольника – среднее арифметическое



Диапазон значений распределения признаков

Рис.6.. Результаты нормализации признаков методами: MinMaxScaler, RobustScaler и StandardScaler

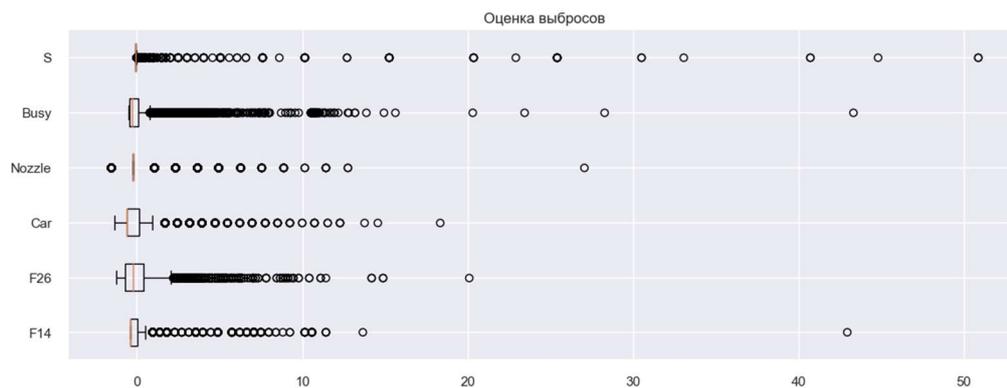


Рис. 7. Диаграммы распределения числовых значений признаков после проведения процедуры нормализации методом Z-масштабирования» StandardScaler

Еще одним важным формальным требованием для разработки модели на основе технологий ML является преобразование качественных признаков исходной статистической выборки $\{N\}$ в числовые значения. Например, признак F6 (Вид населенного пункта) из 8-ми видов, используемых в Федеральной базе данных «Пожары» (далее – ФБД), следует преобразовать до двух: 1 – города; 0 – сельская местность. Виды объектов пожара (признак F12) в ФБД имеет 247 категорий, которые, основываясь на особенности тактики тушения пожаров целесообразно сгруппировать в пять категорий (см. рис.8а).

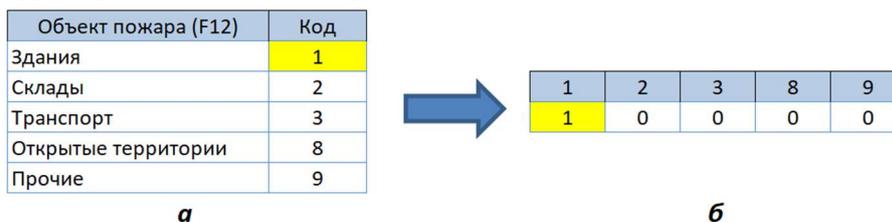


Рис. 8. Схема преобразования качественного (а) признака в числовые значения (б)

Необходимость такого преобразования можно пояснить на следующем примере. В исходной статистической выборке $\{N\}$ для некоторого образца (т.е. пожара) показатель F12 содержал код 1 (здания). Значение 1 или любое другое не имеет какого-либо математического смысла. Например, сложение кода 1 и кода 2 не означает, что объект из категории «здание» перейдет в категорию «транспорт». Поэтому для корректного применения признака F12 необходимо все его пять категорий с помощью унитарного кодирования преобразовать в вектор из пяти признаков. (см. Рис.8б). В результате каждому образцу (т.е. пожару) будет соответствовать значение единицы в соответствующем признаке.

В технологии ML общий вид модели обучения строиться на функции логистической регрессии $f_{w,b}(x)$, которая используется для предсказания неизвестного значения y для данного x ($f_{\hat{w},\hat{b}}(x) \rightarrow y$) вида:

$$f_{w,b}(x) \stackrel{\text{def}}{=} \frac{1}{1 + e^{-(xw+b)}} \quad (2)$$

где w – N -мерный вектор признаков x_i ; b – некоторое действительное число.

Процесс обучения по модели (2) основан на поиске оптимальных значений (\hat{w}, \hat{b}) , при которых модель дает наиболее точные прогнозы. Процедура оптимизации (поиск максимальных значений) осуществляется на основе построения функции правдоподобия:

$$L_{w,b} = \prod_{i=1 \dots N} f_{w,b}(x_i)^{y_i} (1 - f_{w,b}(x_i))^{(1-y_i)}, \quad (3)$$

где Π – арифметическая операция умножения.

Для расчетов удобнее максимизировать логарифм функции правдоподобия ($L_{w,b}$). Это объясняется тем, что данный вид функции является строго возрастающей. Поэтому процедура максимизации функция эквивалентна процедуре максимизации ее аргумента:

$$\ln(L_{w,b}(x)) = \sum_{i=1}^N [y_i \ln f_{w,b}(x) + (1 - y_i) \ln(1 - f_{w,b}(x))]. \quad (4)$$

Для решения задачи поиска максимальных значений применена процедура численной оптимизации по методу градиентного спуска [1,2].

Процесс обучения и тестирования по технологии ML требуется проводить на разных наборах исходных данных, с учетом обучающей выборки. С использованием библиотеки Scikit-learn [5] выполнено разбиение исходного набора данных ($N = 26389$) на две части размером 20% и 80% от его исходного объема (т.е. соответственно 5278 и 21111 образец). Выборка в 20% данных предназначена для тестирования модели.

Проверка качества модели (2) с использованием критерия правдоподобия (4) выполнена с помощью «диаграммы истинности» [6, 7]) (см. Рис.9).

		Прогноз	
		Ординарный пожар	Затяжной пожар
Факт	Ординарный пожар	5226 (99,0 %)	9 (0,2 %)
	Затяжной пожар	25 (0,5 %)	18 (0,3 %)

Рис.2. Матрица запутанности (Confusion Matrix)

Получено, что в 99 % случаев модель (2) по тестовой выборке верно определила характеристики ординарных пожаров. При это следует отметить, что 25 образцов были ложно определены как ординарные.

На рисунке 10 приведены итоговые результаты применения технологий ML по построению распределения признаков и их степень влияния на условия классификации пожаров на «ординарные» и «затяжные». Наиболее значимые факторы, оказывающие влияние на выбор в пользу затяжных пожаров это факторы {F99} и {Busy}. Факторы определяющие ординарные пожары – {F12_8} и {F72_3}.

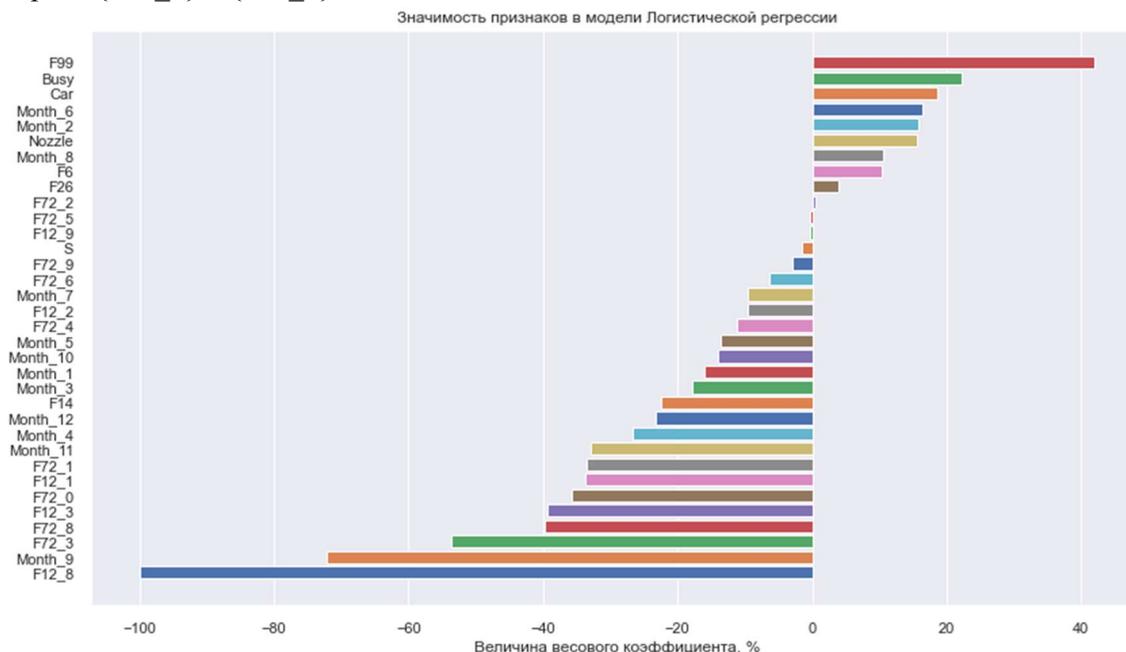


Рис.3. Значимость признаков: F99 – использование СИЗОД; Busy – время занятости на пожаре, мин.; Car – общее количество пожарной техники, ед.; Nozzle – общее количество стволов, ед.; F6 – вид населенного пункта; F26 – расстояние до места пожара, км; S – площадь пожара, м²; F14 – этажность здания; Month_1 ... Month_12 – номер месяца. F72_ – Условия развития пожаров до крупных размеров: 1 - позднее обнаружение пожара; 2 - позднее сообщение о пожаре ; 3 - отсутствие мер по борьбе с пожаром до прибытия подразделений пожарной охраны; 4 - позднее прибытие подразделений пожарной охраны; 5 - поздняя подача огнетушащих средств; 6 - форс-мажорные обстоятельства; 7 - недостатки в организации пожаротушения; 8 - прочие условия. F12_ – Объекты пожара: 1 – здания; 2 – склады; 3 – транспорт; 8 – открытые территории; 9 - прочие

Вывод

Исследование характеристик пожаров представляют научный и практический интерес для изучения в пожарно-спасательных гарнизонах в целях совершенствования деятельности по тушению пожаров. Одним из возможных методов такого исследования является применение технологий машинного обучения, который позволяет, на основе большого объема данных (Big Data), оперативно выявлять факторы и взаимосвязи, оказывающих значительное влияние на процесс пожаротушения с учетом технико-экономических характеристик объектов защиты и природно-климатических особенностей в соответствующем субъекте Российской Федерации.

Список источников

1. Хант Э. Искусственный интеллект. – М.: Мир, 1978. – 558 с.
2. Бурков А. Машинное обучение без лишних слов. — СПб.: Питер, 2020. — 192 с.
3. Харрисон М. Машинное обучение: карманный справочник. Краткое руководство по методам структурированного машинного обучения на Python.: Пер. с англ. - СПб.: ООО «Диалектика», 2020 - 320 с.
4. В.Б. Габдуллин, А.Д. Ищенко Длительность работы звена газодымозащитной службы у очага при тушении затяжных пожаров на объектах энергетики в условиях задымления / Материалы VII Международной научно-практической конференции «Пожаротушение: проблемы, технологии, инновации» в 2 ч. Ч. 2. - М.: Академия ГПС МЧС России, 2020. – 173 с.
5. Репозиторий проекта на сайте GitHub.URL: <https://github.com/VistaSV30/Long.git>. (дата обращения: 21.03.2023)
6. Нолис Ж., Робинсон Э. DataScience для карьериста. – СПб.: Питер, 2021. – 368 с.
7. Спринглер В. Гид по Computer Science. – СПб.: Питер, 2021. – 304 с.

References

1. Hunt E. Artificial Intelligence. – M.: Mir, 1978. – 558 p
2. Burkov A. Machine learning without unnecessary words. - St. Petersburg: Peter, 2020. - 192 p.
3. Harrison M. Machine learning: a pocket guide. A brief guide to structured machine learning methods in Python.: Translated from English - St. Petersburg: Dialectics LLC, 2020 - 320 p.
4. V.B. Gabdullin, A.D. Ishchenko The duration of the work of the gas-smoke protection service at the hearth when extinguishing protracted fires at energy facilities in smoke conditions / Materials of the VII International Scientific and Practical Conference "Firefighting: problems, technologies, innovations" at 2 h. 2. - Moscow: Academy of GPS of the Ministry of Emergency Situations of Russia, 2020. – 173 p .
5. Project repository on GitHub.URL: <https://github.com/VistaSV30/Long.git> . (accessed: 03/21/2023)
6. Nolis J., Robinson E. Data Science for a careerist. – St. Petersburg: Peter, 2021. – 368 p.
7. Sprinkler V. Guide to Computer Science. – St. Petersburg: St. Petersburg, 2021. – 304 p.

Информация об авторах

К.С. Власов - кандидат технических наук

А.А. Порошин - доктор технических наук

Information about the author

K.S. Vlasov - Ph.D. of Engineering Sciences

A.A. Poroshin - Holder of an Advanced Doctorate (Doctor of Science) in Engineering Sciences

Вклад авторов: все авторы сделали эквивалентный вклад в подготовку публикации. Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors: the authors contributed equally to this article. The authors declare no conflicts of interests.

Статья поступила в редакция 10.04.2023; одобрена после рецензирования 08.06.2023; принята к публикации 26.06.2023.

The article was submitted 10.04.2023, approved after reviewing 08.06.2023, accepted for publication 26.06.2023.