

Применение модели логистической регрессии при принятии решений по определению количества привлекаемых сил на ликвидацию лесных пожаров

Дмитрий Валерьевич Медведев , Александр Владимирович Матвеев, Алексей Сергеевич Смирнов

Санкт-Петербургский университет Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий имени Героя Российской Федерации генерала армии Е.Н. Зиничева, г. Санкт-Петербург, Россия

АННОТАЦИЯ

Введение. Прогнозирование количества привлекаемых сил для ликвидации и локализации лесных пожаров является важной и актуальной задачей, оказывающей влияние на эффективность проводимых работ. Однако применение традиционных методов статистического прогнозирования не позволяет получить достоверную оценку целевого показателя в связи с отсутствием ряда признаков при анализе, следствием чего выступает снижение эффективности принимаемых решений.

Цель. Исследование возможности применения модели логистической регрессии для принятия решений о количестве привлекаемых сил на локализацию и ликвидацию лесных пожаров на начальной стадии пожара.


Методы исследования. Применение метода логистической регрессии оценивалось на основе базы данных о лесных пожарах на территории Ленинградской области в период с 2015 по 2023 г., в которой было выделено 16 признаков. Модель логистической регрессии позволяет обучаться на данных, имеющих различные виды распределения, среди которых биномиальное, пуассоновское, Бернулли и другие виды распределения. Математический аппарат, используемый в модели, позволяет оценить апостериорные вероятности для отнесения объектов обучения к соответствующим классам.

Результаты и их обсуждение. Представлены итоги оценки обучения модели в виде матриц ошибок и отчетов о классификации, выполнена визуализация границ решений для случаев использования двух и трех признаков. Результаты показали, что наилучшей точности удалось достичь при использовании всех доступных признаков.

Выводы. Исследование данных лесных пожаров на территории Ленинградской области показало, что присутствуют факторы, которые не учитываются при составлении планов привлечения сил и средств. Применение моделей машинного обучения и, в частности, логистической регрессии, предложенной в данном исследовании, позволяет повысить обоснованность и оперативность принимаемых решений по определению количества привлекаемых сил при лесных пожарах.

Ключевые слова: прогнозирование; машинное обучение; классификация; признаки; алгоритм обучения модели; метрики оценки качества модели; матрица ошибок

Для цитирования: 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/Fire and Explosion Safety. 2024. Т. 33. № 4. С. 84–96. DOI: 10.22227/0869-7493.2024.33.04.84-96

 Медведев Дмитрий Валерьевич, e-mail: meedvedevdv@mail.ru

Application of a logistic regression model in decision-making on determining the number of forces involved in the elimination of forest fires

Dmitriy V. Medvedev , Alexandr V. Matveev, Alexey S. Smirnov

Saint-Petersburg State Fire Service University of the Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters named after the Hero of the Russian Federation, Army General E.N. Zinichev, Saint Petersburg, Russian Federation

ABSTRACT

Introduction. Forecasting the number of involved forces for elimination and localization of forest fires is an important and urgent task that affects the efficiency of the work carried out. However, the use of traditional methods of statistical forecasting does not allow to obtain a reliable assessment of the target indicator, due to the lack of a number of features in the analysis, the consequence of which is a decrease in the effectiveness of decisions.

Objective. Investigation of the possibility of applying the logistic regression model to make decisions on the number of forces to be used for localization and suppression of forest fires at the initial stage of the fire.

Research methods. The application of the logistic regression method was evaluated on the basis of a database of forest fires in the territory of the Leningrad region in the period from 2015 to 2023, in which 16 features were identified. The logistic regression model allows training on data with different types of distribution, including binomial, Poisson, Bernoulli and other types of distribution. The mathematical apparatus used in the model allows us to estimate the posterior probabilities for assigning training objects to the appropriate classes.

Results. The results of the model training evaluation in the form of error matrices and classification reports are presented as results, and visualization of the decision boundaries for the cases of using two and three features is performed. The results show that the best accuracy was achieved using all available features.

Conclusion. The research of forest fires data in the Leningrad region has shown that there are factors that are not taken into account when making plans for the involvement of forces and resources. The use of machine learning models and, in particular, logistic regression, proposed in this study, can improve the validity and efficiency of decisions to determine the number of forces to be involved in forest fires.

Keywords: forecasting; machine learning; classification; signs; model training algorithm; metrics for assessing model quality; error matrix

For citation: Medvedev D.V., Matveev A.V., Smirnov A.S. Application of a logistic regression model in decision-making on determining the number of forces involved in the elimination of forest fires. *Pozharovzryvobezopasnost/ Fire and Explosion Safety*. 2024; 33(4):84-96. DOI: 10.22227/0869-7493.2024.33.04.84-96 (rus).

✉ Dmitriy Valerevich Medvedev, e-mail: meedvedevdv@mail.ru

Введение

Для стран, где леса занимают большую территорию, лесные пожары являются важнейшей угрозой безопасности, потенциально приводящей к огромным ущербам экологии, экономике, а также могут представлять существенную опасность населенным пунктам. В случае реализации данной угрозы оперативному штабу необходимо в минимальные сроки принять решение о расчете привлекаемых сил и средств исходя из существующих планов тушения пожаров лесничеств, а также сводных планов [1]. В случае введения чрезвычайной ситуации или при угрозе населенным пунктам, объектам инфраструктуры к тушению лесных пожаров могут быть привлечены силы и средства МЧС России на основании существующего соглашения в соответствии с Постановлением Правительства РФ¹ и решением комиссии по предупреждению и ликвидации чрезвычайных ситуаций и обеспечению пожарной безопасности.

При этом задача принятия решений по формированию и распределению привлекаемых сил на ликвидацию лесных пожаров в целом является сложной и плохо формализуемой и зачастую принимается в условиях неопределенности [2, 3]. Научное прогнозирование развития лесных пожаров на начальной фазе в целом может потенциально снизить степень неопределенности и обеспечить повышение адекват-

ности принимаемых решений [4, 5]. Развитие лесных пожаров зависит от множества факторов. Накопленная за последние годы статистическая информация о произошедших пожарах может потенциально снизить степень информационной неопределенности на начальном этапе принятия решений [6, 7], однако данные о лесных пожарах достаточно разрежены по причине отсутствия централизованной базы данных, а также неполноты учета факторов. Одним из возможных подходов в данном случае является применение методов машинного обучения, позволяющее компьютерным системам обучаться на основе больших массивов данных и делать на их основе прогнозы для обоснования управленческих решений, в частности при составлении планов привлечения сил и средств, определении условий соглашений с МЧС России для привлечения сил и средств при лесных пожарах. В ряде работ [8, 9] уже применялись методы машинного обучения на статистических данных о лесных пожарах для прогнозирования. По результатам исследования в работе [8] было выделено два метода машинного обучения, показавших наибольшую точность: это модель логистической регрессии и модель повышения градиента. Другие алгоритмы показали более низкие значения точности.

Авторы статьи [8] представили программное обеспечение «Firebird» для определения показателей пожарных рисков и представления соответствующих превентивных мероприятий. По результатам исследования авторы представили интерактивную карту пожарных рисков с возможностью внесения

¹ О привлечении сил и средств федеральных органов исполнительной власти для ликвидации чрезвычайных ситуаций в лесах, возникших вследствие лесных пожаров : Постановление Правительства РФ от 02.12.2017 № 1464.

изменений, базирующуюся на методе опорных векторов и случайного леса. Результаты модели, использующей метод случайного леса, показали более высокую точность, чем модель на опорных векторах. Однако полученные результаты оказались недостаточно точными для предсказания на основе этих моделей. Достичь повышения точности моделей возможно за счет предварительного индексирования показателей пожарного риска. Индексирование показателей осуществляется за счет привлечения группы экспертов, формирующих эвристическую модель. Индекс является мерой количественной оценки риска. Используя характерные для пожаров факторы, в модель вводятся переменные, на основании которых оценивается индекс и проводится анализ пожарного риска. В одной из работ [10] представлена модель по оценке вероятности появления селевого потока как следствия возникновения и развития пожара. В основе модели лежит сопряженная таблица, в которой представлены результаты трех подходов машинного обучения: логистической регрессии, случайных лесов, пороговых значений. Наибольшую точность показал метод случайных лесов.

Несмотря на существующие правила введения чрезвычайных ситуаций в лесах, а также привлечения сил и средств единой государственной системы предупреждения и ликвидации чрезвычайных ситуаций при реагировании на лесные пожары, существуют перспективы повышения обоснованности и оперативности принимаемых решений по определению количества привлекаемых сил и средств при лесных пожарах на основе применения технологий искусственного интеллекта, в частности моделей машинного обучения [11].

Целью настоящей работы является исследование возможности применения модели логистической регрессии для принятия решений о количестве привлекаемых сил на локализацию и ликвидацию лесных пожаров на начальной стадии пожара.

Методы исследования

Статистические данные выступают как основа для машинного обучения [12]. Статистика позволяет осуществлять проверку гипотез и представлять соответствующие оценки. Регрессия — один из методов моделирования и анализа в статистике, позволяющий выявить взаимосвязь между зависимой переменной и набором независимых переменных. В предложенной модели реализуется метод логистической регрессии с применением машинного обучения [13–15]. Существует ряд методов машинного обучения, которые основываются на статистических данных, позволяющих осуществлять классификацию, прогнозирование, кластерный анализ. Поэтому для оценки

количества привлекаемых сил будут предложены несколько вариантов реализации модели логистической регрессии для исследования точности получаемых результатов.

Логистическая регрессия — один из методов математического анализа данных, позволяющих выявить взаимосвязь между целевым признаком и независимыми переменными. Применение машинного обучения для реализации метода логистической регрессии основывается на прогнозировании вероятности наступления категориального события. Реализация данного подхода предполагает классификацию событий в тривиальном случае: 1 — характеризуется как успех, т.е. наступление события, 0 — как неудача. Машинное обучение позволяет осуществить учет необходимого количества факторов в модели без привлечения группы экспертов.

Основные преимущества метода логистической регрессии [16]:

- простота математической реализации при машинном обучении;
- требует меньших ресурсов вычислительной мощности по сравнению с другими методами;
- позволяет наглядно отобразить существующие зависимости между факторами.

Реализация логистической регрессии предполагает выполнение ряда этапов.

Этап 1. Сбор исходных данных

Данные были собраны из открытых источников ИСДМ-Рослесхоз по территории Ленинградской области в период с 2015 по 2023 г., а также использованы данные, предоставленные ГУ МЧС России по Ленинградской области, в которых было выделено 16 признаков (табл. 1) [17].

Этап 2. Предобработка исходных данных

Выделенные переменные возможно классифицировать по типу на две группы: числовые и категориальные.

Метод логистической регрессии в машинном обучении не принимает категориальные признаки в лингвистическом виде, для их кодировки предлагается принимать алгоритм One-hot-encoding [18]. Алгоритм заключается в замене возможных состояний признака на вектор из 0, 1, где 1 характеризуется как наличие данного состояния у признака, 0 — как отсутствие данного состояния. Например, для класса пожарной опасности, соответствующей 1-й категории, вектор будет представлен в следующем виде: 1, 0, 0, 0, 0.

Также важным элементом предобработки данных выступает их нормализация и регуляризация.

Нормализация данных — операция, направленная на исключение ошибок, связанных с раз-

Таблица 1. Переменные, включенные в модель оценки и прогнозирования
Table 1. Variables included in the estimation and forecasting model

Признак Feature	Наименование признака Name of the feature	Описание признака Description of the feature
x_1	Наименование лесничества The name of the forestry	Наименование участкового лесничества на территории Ленинградской области (например, Приозерское) The name of the district forestry in the Leningrad region (for example, Priezerskoye)
x_2	Широта пожара The latitude of the fire	Широта регистрации пожара (например, 58.665000) The latitude of fire registration (for example, 58.665000)
x_3	Долгота пожара The longitude of the fire	Долгота регистрации пожара (например, 30.091944) The longitude of the fire registration (for example, 30.091944)
x_4	Площадь при регистрации, кв. м The area at registration, sq. m	Площадь пожара в момент регистрации (например, 25,0) The area of the fire at the time of registration (for example, 25.0)
x_5	Площадь, пройденная огнем в субъекте РФ, кв. м The area covered by fire in the subject of the Russian Federation, sq. m	Площадь, пройденная огнем в субъекте РФ, без учета площади при регистрации (например, 20,0) The area covered by fire in the subject of the Russian Federation, excluding the area during registration (for example, 20.0)
x_6	Температура воздуха, °C Air temperature, °C	Температура воздуха в градусах Цельсия, измеренная на 9 ч местного времени с ближайшей метеостанции к указанному месту пожара (например, +8,1) The air temperature in degrees Celsius, measured at 9 o'clock local time from the nearest weather station to the specified fire location (for example, +8.1)
x_7	Точка росы, °C Dew point, °C	Точка росы в градусах Цельсия, измеренная на 9 ч местного времени с ближайшей метеостанции к указанному месту пожара (например, +3,0) The dew point in degrees Celsius, measured at 9 o'clock local time from the nearest weather station to the specified fire location (for example, +3.0)
x_8	Суточные осадки, мм Daily precipitation, mm	Суточные осадки в мм, измеренные на 9 ч местного времени с ближайшей метеостанции к указанному месту пожара (например, 33) Daily precipitation in mm, measured at 9 o'clock local time from the nearest weather station to the specified fire (for example, 33)
x_9	Значения показателя класса пожарной природной опасности Values of the fire hazard class indicator	Класс пожарной природной опасности на основе влажности почвенного покрова, рассчитываемый в соответствии с методиками расчета класса пожарной природной опасности, критерием для сброса или понижения класса пожарной природной опасности на утро текущего дня является количество суточных осадков на 9 ч утра местного времени (осадки за вчерашний день + ночь) (например, 1463) The class of fire natural hazard based on the moisture content of the ground cover, calculated in accordance with the methods of calculating the class of fire natural hazard, the criterion for dumping or lowering the class of fire natural hazard in the morning of the current day is the amount of daily precipitation for 9 am local time in the morning (precipitation for yesterday + night), (for example, 1,463)
x_{10}	Класс пожарной опасности Fire hazard class	Класс пожарной опасности в соответствии со значением показателя класса пожарной природной опасности (например, 3) Fire hazard class according to the value of the attribute, the value of the fire hazard class indicator (for example, 3)
x_{11}	Расстояние до пожарной части, м The distance to the fire station, m	Евклидово расстояние от точки регистрации пожара до координат ближайшей пожарной части (например, 1500) The Euclidean distance from the fire registration point to the coordinates of the nearest fire station (for example, 1,500)
x_{12}	Широта пожарной части The latitude of the fire station	Широта расположения пожарной части (например, 58.738500) The latitude of the location of the fire station (for example, 58.738500)
x_{13}	Долгота пожарной части The longitude of the fire station	Долгота расположения пожарной части (например, 29.849600) The longitude of the location of the fire station (for example, 29.849600)

Признак Feature	Наименование признака Name of the feature	Описание признака Description of the feature
x_{14}	Количество сил пожарной части, чел. The number of fire department forces, people	Количество сил, находящихся в боевом расчете пожарной части (например, 15) The number of forces in the combat calculation of the fire department (for example, 15)
x_{15}	Количество техники пожарной части The number of fire department equipment	Количество техники, находящейся в боевом расчете пожарной части (например, 3) The number of vehicles in the combat calculation of the fire department (for example, 3)
x_{16}	Количество пожарно-технического вооружения пожарной части The number of fire-technical weapons of the fire department	Количество пожарно-технического вооружения, находящегося в боевом расчете пожарной части (например, 8) The number of fire-technical weapons in the combat calculation of the fire department (for example, 8)

личным физическим смыслом признаков, а также их масштабом. Нормализация позволяет привести признаки к единому масштабу значений, что приведет к более эффективной обработке модели. Одним из методов нормализации является использование математического ожидания и дисперсии.

Регуляризация данных — операция, направленная на уменьшение абсолютных значений весов по каждому из признаков. В случае, если модель обладает свойством мультиколлинеарности, то решений становится бесконечно много, так как веса признаков могут принимать континуальное множество решений [19]. Следствием чего выступает переобучение модели, выявление слишком сложных зависимостей. Предлагается 3 подхода к регуляризации весов: l_1 — регуляризация (LASSO), l_2 — регуляризация (Ridge), Elastic net — комбинация l_1 и l_2 [20].

l_1 — регуляризация штрафует модель на сумму модулей всех весов, так как функция модуля не дифференцируема в нуле, это накладывает ограничения в виде возможного зануления весов для некоторых признаков, а также отсутствия аналитического решения.

l_2 — регуляризация штрафует модель на сумму квадратов ее весов, квадратичная функция является дифференцируемой, что позволяет получить аналитическое решение.

Этап 3. Обучение модели

Данный этап предполагает проведение обучения модели логистической регрессии на обучающем наборе с целью получения оценок по выбранным метрикам для первичной оценки качества модели. В случае получения неудовлетворительных результатов необходимо вернуться на предыдущие этапы для корректировки данных.

Этап 4. Апробирование модели на тестовом наборе данных

Данный этап позволяет на обученной модели провести оценку согласно метрикам на отложенном наборе данных. В результате оценки принимается решение об удовлетворительности применяемой модели для целевой переменной. В случае получения неудовлетворительных результатов предполагается вернуться на предыдущий этап для корректировки гиперпараметров модели.

Теоретические основы

Признаки, используемые в модели, основываются на исходном наборе предобработанных данных, для которых возможно применение методов статистического машинного обучения. Распределение вероятностей пожароопасной ситуации носит дискретный характер, для представления которых возможно использовать биномиальное, пуассоновское, Бернулли и иные виды распределения. Для понимания работы алгоритма рассмотрим биномиальный закон распределения, для которого характерна определенная степень постоянства вероятности. Согласно закону биномиального распределения, случайная величина Y определяется как:

$$P(Y = y) = \binom{n}{y} \cdot p^y \cdot (1 - p)^{n-y}, \quad y = 0, 1, \dots, n, \quad (1)$$

где Y — случайная величина;

n — число испытаний;

p — вероятность наступления события.

Математическое ожидание биномиального закона распределения определяется как произведение общего числа испытаний на вероятность наступления события $M(Y) = np$, среднеквадратичное отклонение определяется как: $\sigma = \sqrt{np(1-p)}$. В простейшем случае целевая переменная может иметь бинарное представ-

ление, определяющее: 0 — силы не привлекаются, 1 — силы привлекаются. Таким образом, логит-модель регрессии будет представлена как:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon, \quad (2)$$

где $\log(p/1-p)$ — логистическая функция;

p — вероятность привлечения сил;

β_i — коэффициенты, отражающие влияние переменных на результат прогностической модели;

x_i — независимые переменные, характеризующие объект;

ε — аддитивная случайная ошибка, имеющая стандартное логистическое распределение. Возведя в степень в обе части логит-модели, возможно перейти к следующей форме представления модели:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}}, \quad (3)$$

где β_i — оценка коэффициентов модели при условии, что e стремится к минимуму.

Определение коэффициентов логистической регрессии на практике зачастую реализуют методом максимального правдоподобия. В основе метода лежит функция правдоподобия, характеризующая плотность вероятности целевой переменной по выборке. Задачей является поиск параметров модели, при которых заданная функция правдоподобия принимает максимальное значение. На практике максимизируют не саму функцию, а натуральный логарифм функции правдоподобия, так как максимум в обоих случаях достигается при равных значениях искомых параметров.

Пусть P_i — вероятность наступления ожидаемого события, т.е. “1”, тогда:

$$P_i = \text{Prob}(Y_i = 1). \quad (4)$$

Значение вероятности зависит от $x_i w$, где x_i — строка матрицы независимых переменных, w — вектор коэффициентов регрессии.

Таким образом, функция правдоподобия будет равна:

$$\begin{aligned} L^* &= \sum_{i \in I_1} \ln P_i(x_i) + \sum_{i \in I_0} \ln(1 - P_i(x_i)) = \\ &= \sum_{i=1}^k (Y_i \cdot \ln P_i(x_i) + (1 - Y_i) \cdot \ln(1 - P_i(x_i))), \end{aligned} \quad (5)$$

где I_0, I_1 — множество исходных данных, для которых $Y_i = 0, I_i = 1$ соответственно.

Основным отличием логистической регрессии от линейной регрессии является вид предсказания результата, который является вероятностной функ-

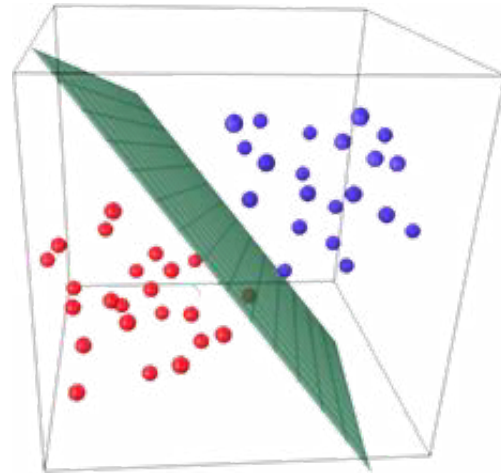


Рис. 1. Разделяющая граница классов для модели логистической регрессии при двух переменных

Fig. 1. Dividing class boundary for a logistic regression model with two variables

цией принадлежности к заданным классам. Результат логистической регрессии находится в интервале 0; 1. Вся область исходных данных разделяется границей, разделяющей классы. В тривиальном случае для одной независимой переменной разделяющей границей является прямая, в случае двух переменных — плоскость (рис. 1), для более сложных случаев — гиперплоскость.

Разделяющая плоскость на рис. 1 является линейным дискриминантом, так как является линейной по отношению к построенной функции.

Логит-модель регрессии может быть реализована с применением нейронных сетей и применяться для оценки и прогнозирования пожароопасных ситуаций (рис. 2). Исходными данными для реализации будет выступать кортеж переменных (Y, X_i) , весами нейронной сети — коэффициенты логистической регрессии. При реализации модели машинного обу-

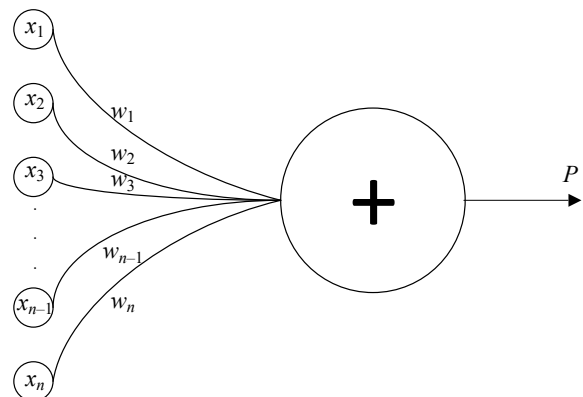


Рис. 2. Логистическая регрессия в структурном виде нейронной сети

Fig. 2. Logistic regression in the structural form of a neural network

Таблица 2. Оценка качества классификации
Table 2. Classification quality assessment

Модель Model	Фактически положительные Actually positive	Фактически отрицательные Actually negative
Положительные Positive	TP	FP
Отрицательные Negative	FN	TN

TP — true positive, правильно классифицированные ожидаемые события.

TN — true negatives, правильно классифицированные не ожидаемые события.

FN — false negatives, неправильно классифицированные ожидаемые события.

FP — false positive, неправильно классифицированные не ожидаемые события.

TP — true positive, correctly classified expected events.

TN — true negatives, correctly classified unexpected events.

FN — false negatives, incorrectly classified expected events.

FP — false positive, incorrectly classified unexpected events.

чения результатом может являться 0 (силы не привлекаются) и 1 (силы привлекаются).

Оценку качества модели логистической регрессии предлагается осуществить с помощью стандартных метрик библиотеки Sklearn языка программирования Python. Также достаточно часто используется ROC-кривая для оценки результатов, характеризующая взаимосвязь между количеством правильно классифицированных ожидаемых событий (т.е. 1) и количеством неправильно классифицированных неожиданных событий (т.е. 0). Данные, характеризующие варианты классификации набора данных, представлены в табл. 2 [10, 21].

На основании этих показателей определяются значения двух критериев для оценки: точность (отражающая специфичность модели) и полнота (отражающая чувствительность модели).

Точность (S_p) — число правильно классифицированных ожидаемых событий к общему числу классификаций ожидаемых событий:

$$S_p = \frac{TP}{TP + FP}. \tag{6}$$

Полнота (S_{recall}) — число правильно классифицированных ожидаемых событий к общему числу элементов данного класса (фактически положительных):

$$S_{recall} = \frac{TP}{TP + FN}. \tag{7}$$

Таким образом, высокое значение показателя S_{recall} представляет верный результат при наличии ожидаемых событий (обнаруживает ожидаемые события).

В процессе подготовки данных для машинного обучения необходимо разделить набор на две составляющие: обучающий и тестовый. Для реализации был использован язык программирования Python.

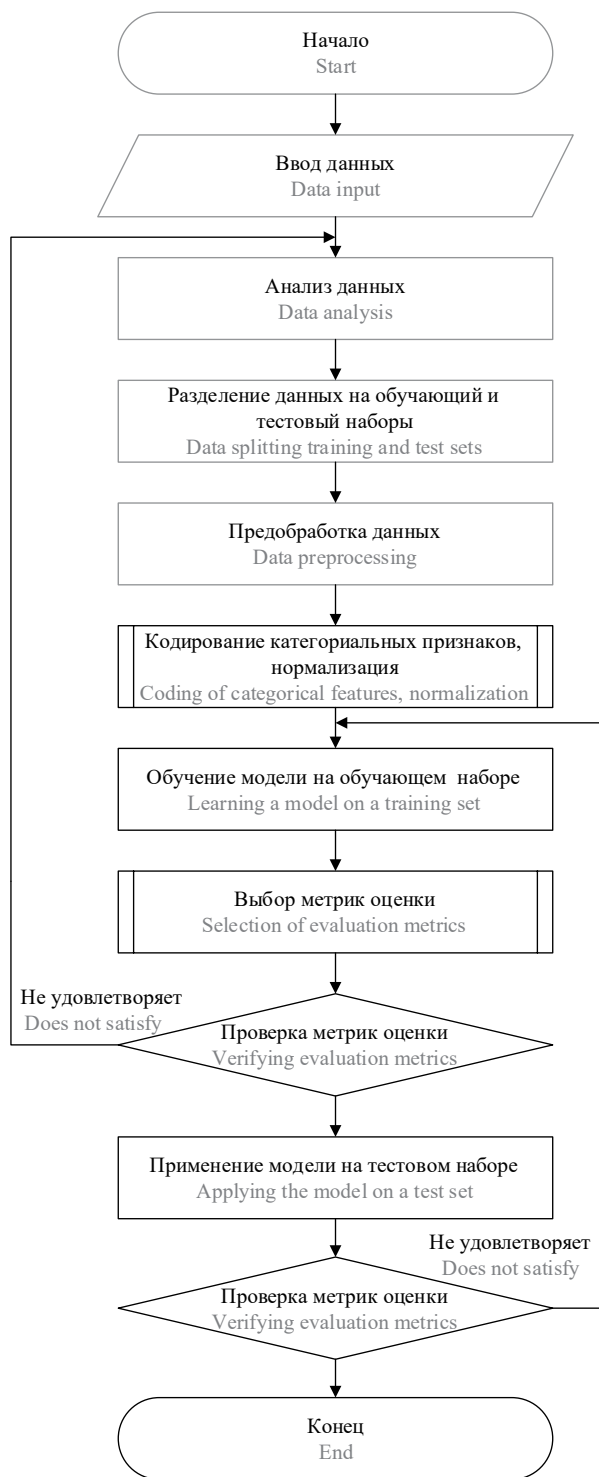


Рис. 3. Блок-схема алгоритма логистической регрессии
Fig. 3. Block diagram of the logistic regression algorithm

Блок-схема алгоритма логистической регрессии представлена на рис. 3.

Результаты и их обсуждение

Поиск наилучших параметров для модели логистической регрессии осуществлялся по сетке с помощью встроенного в библиотеку Sklearn модуля GridSearchCV, в качестве метрики для оценки исполь-

зовался отчет о классификации и матрица ошибок. База данных содержала 376 объектов, на которых необходимо провести обучение и оценку.

Целевая переменная была разделена на три класса по количеству привлеченных сил: не более 10 человек — 1; от 10 до 24 человек — 2; более 24 человек — 0.

На рис. 4 представлена гистограмма распределения количества экземпляров каждого из классов целевой переменной.

Так как исходные данные имеют незначительное количество экземпляров миноритарного класса, было проведено синтетическое увеличение числа объектов данного класса с использованием алгоритма RandomOverSampler библиотеки imblearn [22]. В результате балансировки распределения объектов по подклассам получилось 732 экземпляра.

Количество независимых переменных (в данном случае 16) не позволяет визуализировать результаты классификации, поэтому было принято решение применить метод главных компонент для уменьшения размерности до 2 и 3 признаков. На рис. 5 и 6 представлены границы решений классификации при

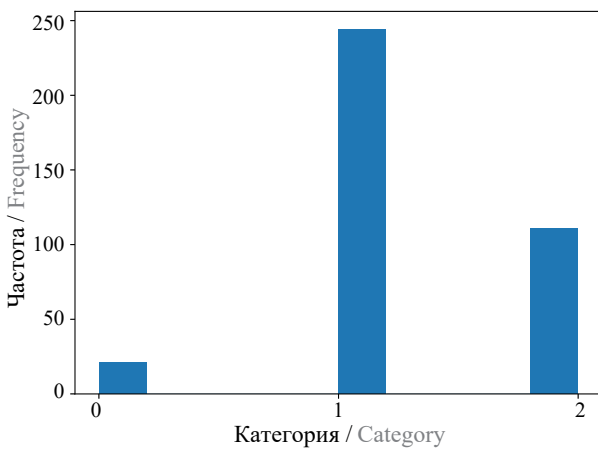


Рис. 4. Гистограмма распределения экземпляров по классам
Fig. 4. Histogram of distribution of specimens by classes

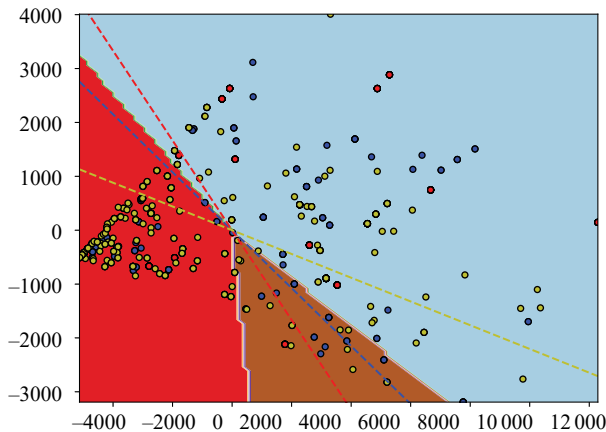


Рис. 5. Границы решений для схемы ovr
Fig. 5. Decision boundary of LogisticRegression with ovr

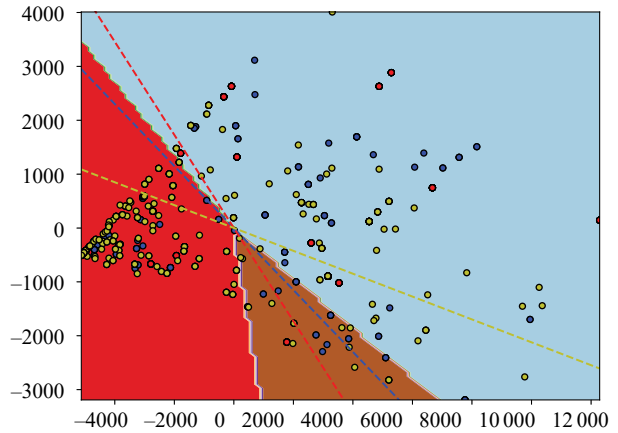


Рис. 6. Границы решений для схемы multinomial
Fig. 6. Decision boundary of Logistic Regression with multinomial

уменьшении размерности до 2 признаков для схемы классификации «один класс против остальных (ovr)» и схемы, использующей перекрестную энтропийную потерю (multinomial) соответственно. Отчет о результатах классификации представлен в табл. 3.

На рис. 7 приведена визуализация решений классификации при уменьшении размерности до 3

Таблица 3. Отчет о результатах классификации при использовании 2 признаков
Table 3. Report on classification results using 2 features

Классы Classes	Метрики Metrics			
	Отчет о классификации Classification Report is			
	Точность Accuracy	Полнота Fullness	f1-оценка f1-score	Количество объектов класса Number of class objects
0	0,52	0,71	0,60	244
1	0,47	0,68	0,55	244
2	0,43	0,08	0,13	244
Точность Accuracy	Не измеряется Not measured	Не измеряется Not measured	0,49	732
Макро-усреднение Macro averaging	0,47	0,49	0,43	723
Микро-усреднение Micro averaging	0,47	0,49	0,43	723

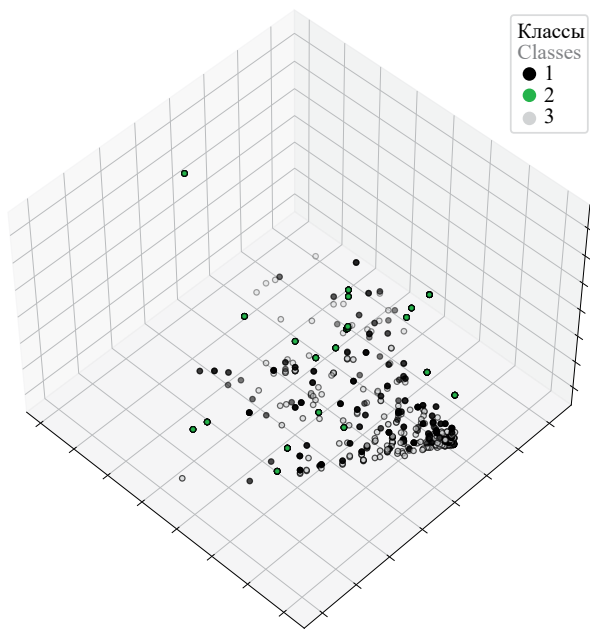


Рис. 7. Визуализация решений классификации при использовании 3 признаков
Fig. 7. Visualization of classification decisions using 3 features

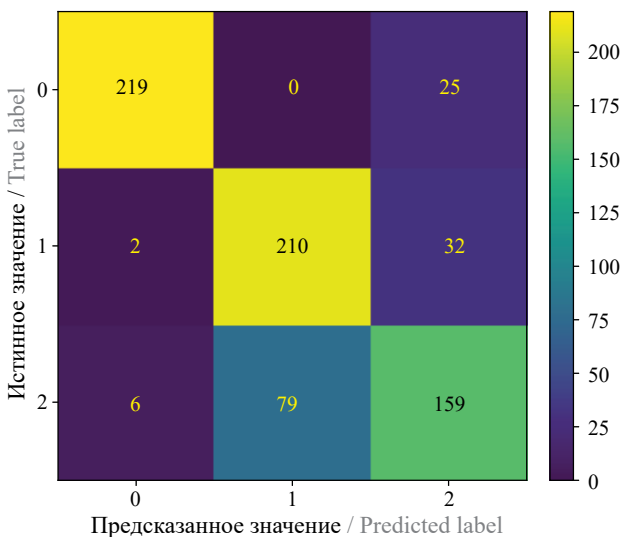


Рис. 8. Матрица ошибок классификации при использовании 3 признаков
Fig. 8. Classification error matrix when using 3 features

признаков. На рис. 8 представлена матрица ошибок при использовании 3 признаков. Отчет о результатах классификации представлен в табл. 4.

На рис. 9 представлена матрица ошибок при использовании всех признаков. Отчет о результатах классификации представлен в табл. 5.

Таким образом, уменьшение размерности позволило визуализировать полученные результаты классификации, однако при использовании двух признаков точность классификации оказалась самой низкой, а в случае использования трех признаков результат достаточно близок к случаю, при котором использовались все доступные для анализа признаки.

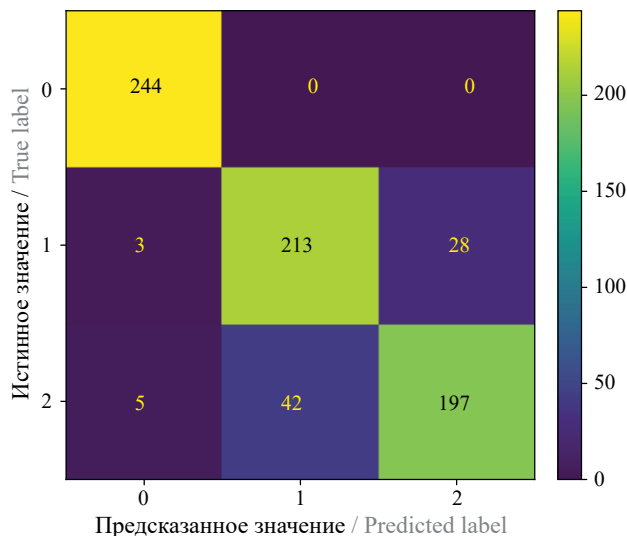


Рис. 9. Матрица ошибок классификации при использовании всех признаков
Fig. 9. Matrix of classification errors when using all the features

Таблица 4. Отчет о результатах классификации при использовании 3 признаков
Table 4. Report on classification results using 3 features

Классы Classes	Метрики Metrics			
	Отчет о классификации Classification report			
	Точность Accuracy	Полнота Fullness	f1-оценка f1-score	Количество объектов класса Number of class objects
0	0,96	0,90	93	244
1	0,73	0,86	0,79	244
2	0,74	0,65	0,69	244
Точность Accuracy	Не измеряется Not measured	Не измеряется Not measured	0,80	732
Макро-усреднение Macro averaging	0,1	0,80	0,80	723
Микро-усреднение Micro averaging	0,81	0,80	0,80	723

Таблица 5. Отчет о результатах классификации при использовании всех признаков**Table 5.** Report on the classification results using all the features

Классы Classes	Метрики Metrics			
	Отчет о классификации Classification report			
	Точность Accuracy	Полнота Fullness	f1- оценка f1-score	Количество объектов класса Number of class objects
0	0,97	1,00	0,98	244
1	0,84	0,87	0,85	244
2	0,88	0,81	0,84	244
Точность Accuracy	Не измеряется Not measured	Не измеряется Not measured	0,89	732
Макро- усреднение Macro averaging	0,89	0,89	0,89	723
Микро- усреднение Micro averaging	0,81	0,89	0,89	723

Выводы

Уровень существующих угроз и значимость возможных последствий лесных пожаров актуализируют необходимость оперативной и достоверной оценки возможного развития ситуации с целью выработки адекватных управленческих решений.

В статье на основе использования данных о 16 первичных признаках, собранных на территории Ленинградской области за период с 2015 по 2023 г., исследованы возможности применения модели логистической регрессии для определения количества привлекаемых сил и средств МЧС России на ликвидацию лесных пожаров.

Результат реализации модели логистической регрессии показали наилучшие результаты по используемым метрикам в случае, когда использовались все доступные признаки, что подтверждает гипотезу о необходимости учета множества факторов при принятии решений при лесных пожарах. Однако существенным недостатком является отсутствие возможности визуализации полученных результатов, что затрудняет процесс принятия решения ответственными лицами.

Применение моделей машинного обучения и, в частности, логистической регрессии, предложенной в данном исследовании, позволяет повысить обоснованность и оперативность принимаемых решений по определению количества привлекаемых сил при лесных пожарах.

СПИСОК ИСТОЧНИКОВ

1. Вилков В.Б., Горшкова Е.Е., Черных А.К. Решение задачи нахождения оптимального маршрута патрулирования действующих лесных пожаров в заданном районе // Научно-аналитический журнал «Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России». 2021. № 3. С. 90–98. EDN ZYOYFF.
2. Молчанов А.В. Концептуальная модель формирования рационального состава группировки сил различной ведомственной принадлежности при ликвидации чрезвычайных ситуаций в лесах регионального характера, возникших вследствие лесных пожаров // Научные и образовательные проблемы гражданской защиты. 2015. № 4 (27). С. 95–102. EDN VCITFJ.
3. Карапузиков А.А., Дьяков В.Ф., Кокишаров А.В., Дьяков М.В., Ставриниди С.Ю., Белкин Д.С. К вопросу об управлении силами и средствами при тушении лесных пожаров // Техносферная безопасность. 2020. № 2 (27). С. 16–27. EDN URZSEG.
4. Матвеев А.В., Богданова Е.М. Классификация методов прогнозирования чрезвычайных ситуаций // Национальная безопасность и стратегическое планирование. 2018. № 4 (24). С. 61–70. EDN YTPZYL.
5. Станкевич Т.С. Прогнозирование пространственного поведения лесного пожара при неопределенности и нестационарности процесса // Известия высших учебных заведений. Лесной журнал. 2021. № 1 (379). С. 20–34. DOI: 10.37482/0536-1036-2021-1-20-34. EDN YTFJYI.
6. Станкевич Т.С. Разработка метода оперативного прогнозирования динамики развития лесного пожара посредством искусственного интеллекта и глубокого машинного обучения // Вестник Иркутского государственного технического университета. 2018. Т. 22. № 9 (140). С. 111–120. DOI: 10.21285/1814-3520-2018-9-111-120. EDN YGGSLR.
7. 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.

8. Madaio M., Chen S.T., Haimson O., Zhang W., Cheng X., Hinds-Aldrich M. et al. Firebird: Predicting fire risk and prioritizing fire inspections in Atlanta // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. Pp. 185–194. DOI: 10.1145/2939672.2939682
9. Nikolopoulos E.I., Destro E., Bhuiyan M., Borga M., Anagnostou E. Evaluation of predictive models for post-fire debris flow occurrence in the western United States // Natural Hazards and Earth System Sciences. 2018. Vol. 18. No. 9. Pp. 2331–2343. DOI: 10.5194/nhess-18-2331-2018
10. Pham B.T., Jaafari A., Avand M., Al-Ansari N., Dinh Du T., Yen H.P. et al. Performance evaluation of machine learning methods for forest fire modeling and prediction // Symmetry. 2020. Vol. 12. No. 6. P. 1022. DOI: 10.3390/sym12061022
11. Максимов А.В. Методы поддержки принятия решений в оперативном управлении при чрезвычайных ситуациях: обзор исследований // Национальная безопасность и стратегическое планирование. 2023. № 2 (42). С. 91–102. DOI: 10.37468/2307-1400-2023-2-91-102. EDN CJCPWN.
12. Бутырский Е.Ю., Мамвеев А.В. Математическое моделирование систем и процессов. СПб.: Информационный издательский учебно-научный центр «Стратегия будущего», 2022. 733 с. DOI: 10.37468/book_011222. EDN CCRIRT.
13. Ray S. A quick review of machine learning algorithms // 2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE, 2019. Pp. 35–39. DOI: 10.1109/COMITCon.2019.8862451
14. Abid F. A survey of machine learning algorithms based forest fires prediction and detection systems // Fire technology. 2021. Vol. 57. No. 2. Pp. 559–590. DOI: 10.1007/s10694-020-01056-z
15. Sayad Y.O., Mousannif H., Al Moatassime H. Predictive modeling of wildfires: A new dataset and machine learning approach // Fire safety journal. 2019. Vol. 104. Pp. 130–146. DOI: 10.1016/j.firesaf.2019.01.006
16. Yang Y., Loog M. A benchmark and comparison of active learning for logistic regression // Pattern Recognition. 2018. Vol. 83. Pp. 401–415. DOI: 10.1016/j.patcog.2018.06.004
17. Медведев Д.В. Модель прогнозирования лесных пожаров на основе нейро-нечеткой системы ANFIS // Научно-аналитический журнал «Вестник Санкт-Петербургского университета Государственной противопожарной службы МЧС России». 2023. № 4. С. 185–198. DOI: 10.61260/2218-130X-2024-2023-4-185-198. EDN MXLKB1.
18. Karthiga R., Usha G., Raju N., Narasimhan K. Transfer learning based breast cancer classification using one-hot encoding technique // 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). IEEE, 2021. Pp. 115–120. DOI: 10.1109/ICAIS50930.2021.9395930
19. Salehi F., Abbasi E., Hassibi B. The impact of regularization on high-dimensional logistic regression // Advances in Neural Information Processing Systems. 2019. Vol. 32.
20. Liang X., Jacobucci R. Regularized structural equation modeling to detect measurement bias: Evaluation of lasso, adaptive lasso, and elastic net // Structural Equation Modeling: a Multidisciplinary Journal. 2020. Vol. 27. No. 5. Pp. 722–734. DOI: 10.1080/10705511.2019.1693273
21. Molovtsev M.D., Sineva I.S. Classification algorithms analysis in the forest fire detection problem // 2019 International Conference “Quality Management, Transport and Information Security, Information Technologies” (IT&QM&IS). IEEE, 2019. Pp. 548–553. DOI: 10.1109/ITQMIS.2019.8928398
22. Hayaty M., Muthmainah S., Ghufuran S.M. Random and synthetic over-sampling approach to resolve data imbalance in classification // International Journal of Artificial Intelligence Research. 2020. Vol. 4. No. 2. Pp. 86–94. DOI: 10.29099/ijair.v4i2.152

REFERENCES

1. Vilkov V.B., Gorshkova E.E., Chernykh A.K. Solving the problem of finding the optimal route for patrolling active forest fires in a given area. *Scientific and analytical journal “Bulletin of the St. Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia”*. 2021; 3:90-98. EDN ZYOYFF. (rus).
2. Molchanov A.V. Conceptual model of formation of rational structure of group of forces of various departmental accessory, at elimination of emergency situations in the woods of regional character which arose owing to forest fires. *Scientific and Educational Problems of Civil Protection*. 2015; 4(27):95-102. EDN VCITFJ. (rus).
3. Karapuzikov A.A., Dyakov V.F., Koksharov A.V., Dyakov M.V., Stavrinidi S.Y., Belkin D.S. On the issue of managing forces and means for extinguishing forest fires. *Technospheric Safety*. 2020; 2(27):16-27. EDN URZSEG. (rus).
4. Matveev A.V., Bogdanova E.M. The classification for the methods of prediction emergency situations. *National Security and Strategic Planning*. 2018; 4(24):61-70. EDN YTPZYL. (rus).
5. Stankevich T.S. Forecasting the spatial behavior of a forest fire at uncertainty and instability of the process. *News of Higher Educational Institutions. Forest Magazine*. 2021; 1(379):20-34. DOI: 10.37482/0536-1036-2021-1-20-34. (rus).

6. Stankevich T.S. Development of operational prediction method of forest fire dynamics based on artificial intelligence and deep machine learning. *Bulletin of the Irkutsk State Technical University*. 2018; 22(9):111-120. DOI: 10.21285/1814-3520-2018-9-111-120 (rus).
7. Matveev A.V., Matiev R.T. Decision-making during fires in mountainous areas : a comparative analysis of monitoring methods. *National Security and Strategic Planning*. 2023; 2(42):76-90. DOI: 10.37468/2307-1400-2023-2-76-90 (rus).
8. Madaio M., Chen S.T., Haimson O., Zhang W., Cheng X., Hinds-Aldrich M. et al. Firebird: Predicting fire risk and prioritizing fire inspections in Atlanta. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016; 185-194. DOI: 10.1145/2939672.2939682
9. Nikolopoulos E.I., Destro E., Bhuiyan M., Borga M., Anagnostou E. Evaluation of predictive models for post-fire debris flow occurrence in the western United States. *Natural Hazards and Earth System Sciences*. 2018; 18(9):2331-2343. DOI: 10.5194/nhess-18-2331-2018
10. Pham B.T., Jaafari A., Avand M., Al-Ansari N., Dinh Du T., Yen H.P. et al. Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry*. 2020; 12(6):1022. DOI: 10.3390/sym12061022
11. Maximov A.V. Decision support methods in emergency management: a review of research. *National Security and Strategic Planning*. 2023; 2(42):91-102. DOI: 10.37468/2307-1400-2023-2-91-102 (rus).
12. Butyrsky E.Yu., Matveev A.V. *Mathematical modeling of systems and processes*. 2022; 733. DOI 10.37468/book_011222 (rus).
13. Ray S. A quick review of machine learning algorithms. *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)*. 2019; 35-39. DOI: 10.1109/COMITCon.2019.8862451
14. Abid F. A survey of machine learning algorithms based forest fires prediction and detection systems. *Fire Technology*. 2021; 57(2):559-590. DOI: 10.1007/s10694-020-01056-z
15. Sayad Y.O., Mousannif H., Al Moatassime H. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*. 2019; 104:130-146. DOI: 10.1016/j.firesaf.2019.01.006
16. Yang Y., Loog M. A benchmark and comparison of active learning for logistic regression. *Pattern Recognition*. 2018; 83:401-415. DOI: 10.1016/j.patcog.2018.06.004
17. Medvedev D.V. Implementation of the ANFIS neuro-fuzzy system for forest fire management. *Scientific and Analytical Journal "Bulletin of the St. Petersburg University of the State Fire Service of the Ministry of Emergency Situations of Russia"*. 2023; 4:185-198. DOI: 10.61260/2218-130X-2024-2023-4-185-198 (rus).
18. Karthiga R., Usha G., Raju N., Narasimhan K. Transfer learning based breast cancer classification using one-hot encoding technique. *2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS)*. 2021; 115-120. DOI: 10.1109/ICAIS50930.2021.9395930
19. Salehi F., Abbasi E., Hassibi B. The impact of regularization on high-dimensional logistic regression. *Advances in Neural Information Processing Systems*. 2019; 32.
20. Liang X., Jacobucci R. Regularized structural equation modeling to detect measurement bias: Evaluation of lasso, adaptive lasso, and elastic net. *Structural Equation Modeling : a Multidisciplinary Journal*. 2020; 27(5):722-734. DOI: 10.1080/10705511.2019.1693273
21. Molovtsev M.D., Sineva I.S. Classification algorithms analysis in the forest fire detection problem. *2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*. IEEE, 2019; 548-553. DOI: 10.1109/ITQMIS.2019.8928398
22. Hayaty M., Muthmainah S., Ghufuran S.M. Random and synthetic over-sampling approach to resolve data imbalance in classification. *International Journal of Artificial Intelligence Research*. 2020; 4(2):86-94. DOI: 10.29099/ijair.v4i2.152

Поступила 11.05.2024, после доработки 12.06.2024;

принята к публикации 28.06.2024

Received May 11, 2024; Received in revised form June 12, 2024;

Accepted June 28, 2024

Информация об авторах

МЕДВЕДЕВ Дмитрий Валерьевич, адъюнкт, Санкт-Петербургский университет Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий имени Героя Российской Федерации генерала армии Е.Н. Зиничева, Россия, 196105, г. Санкт-Петербург, Московский пр-т, 149; SPIN-код: 8464-6705; Scopus AuthorID: 57197819511; ORCID: 0009-0002-9436-4376; e-mail: meedvedevdv@mail.ru

Information about the authors

Dmitriy V. MEDVEDEV, Graduate, Saint-Petersburg State Fire Service University of the Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters named after the Hero of the Russian Federation, Army General E.N. Zinichev, Moskovskiy Ave., 149, Saint-Petersburg, 196105, Russian Federation; SPIN-code: 8464-6705; Scopus AuthorID: 57197819511; ORCID: 0009-0002-9436-4376; e-mail: meedvedevdv@mail.ru

МАТВЕЕВ Александр Владимирович, канд. техн. наук, доцент, заведующий кафедрой прикладной математики и информационных технологий, Санкт-Петербургский университет Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий имени Героя Российской Федерации генерала армии Е.Н. Зиничева, Россия, 196105, г. Санкт-Петербург, Московский пр-т, 149; SPIN-код: 5778-8832; Scopus AuthorID: 57197819511; ORCID: 0000-0002-0778-3218; e-mail: fcvega_10@mail.ru

СМИРНОВ Алексей Сергеевич, д-р техн. наук, профессор, первый заместитель начальника, Санкт-Петербургский университет Государственной противопожарной службы Министерства Российской Федерации по делам гражданской обороны, чрезвычайным ситуациям и ликвидации последствий стихийных бедствий имени Героя Российской Федерации генерала армии Е.Н. Зиничева, Россия, 196105, г. Санкт-Петербург, Московский пр-т, 149; SPIN-код: 1677-1402; ORCID: 0000-0003-1661-9089; e-mail: sas@igps.ru

Вклад авторов:

Медведев Д.В. — сбор материала; разработка методик и проведение эксперимента; получение, обработка и анализ экспериментальных данных; написание статьи.

Матвеев А.В. — сбор материала; разработка методик и проведение эксперимента; написание статьи; научное редактирование текста.

Смирнов А.С. — идея; научное редактирование текста. Авторы заявляют об отсутствии конфликта интересов.

Alexander V. MATVEEV, Cand. Sci. (Eng.), Assistant Professor, Head of Department of Applied Mathematics and Information Technologies, Saint-Petersburg State Fire Service University of the Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters named after the Hero of the Russian Federation, Army General E.N. Zinichev, Moskovskiy Ave., 149, Saint-Petersburg, 196105, Russian Federation; SPIN-code: 5778-8832; Scopus AuthorID: 57197819511; ORCID: 0000-0002-0778-3218; e-mail: fcvega_10@mail.ru

Alexey S. SMIRNOV, Dr. Sci. (Eng.), Professor, First Deputy Head, Saint-Petersburg State Fire Service University of the Ministry of the Russian Federation for Civil Defense, Emergencies and Elimination of Consequences of Natural Disasters named after the Hero of the Russian Federation, Army General E.N. Zinichev, Moskovskiy Ave., 149, Saint-Petersburg, 196105, Russian Federation; SPIN-code: 1677-1402; ORCID: 0000-0003-1661-9089; e-mail: sas@igps.ru

Contribution of the authors:

Medvedev D.V. — collection of material; development of methodology and experiment; acquisition, processing and analysis of experimental data; writing of article.

Matveev A.V. — collection of material; development of methodology and experiment; acquisition; writing of article; scientific editing of the text.

Smirnov A.S. — idea; scientific editing of the text.

The authors declare that there is no conflict of interest.