**Затяжные пожары**

В 2021 году в Магаданской, Московской областях и Забайкальском крае произошло соответственно 877, 20520 и 4992-а пожара для тушения, которых привлеклось 1472, 38439 и 7471-а ед. пожарной техники. Количество пожаров и техники, как и большинство других показателей характеризующих процесс тушения, и интервал времени занятости пожарных подразделений на пожаре в том числе (Рисунок 1), являются случайными величинами.

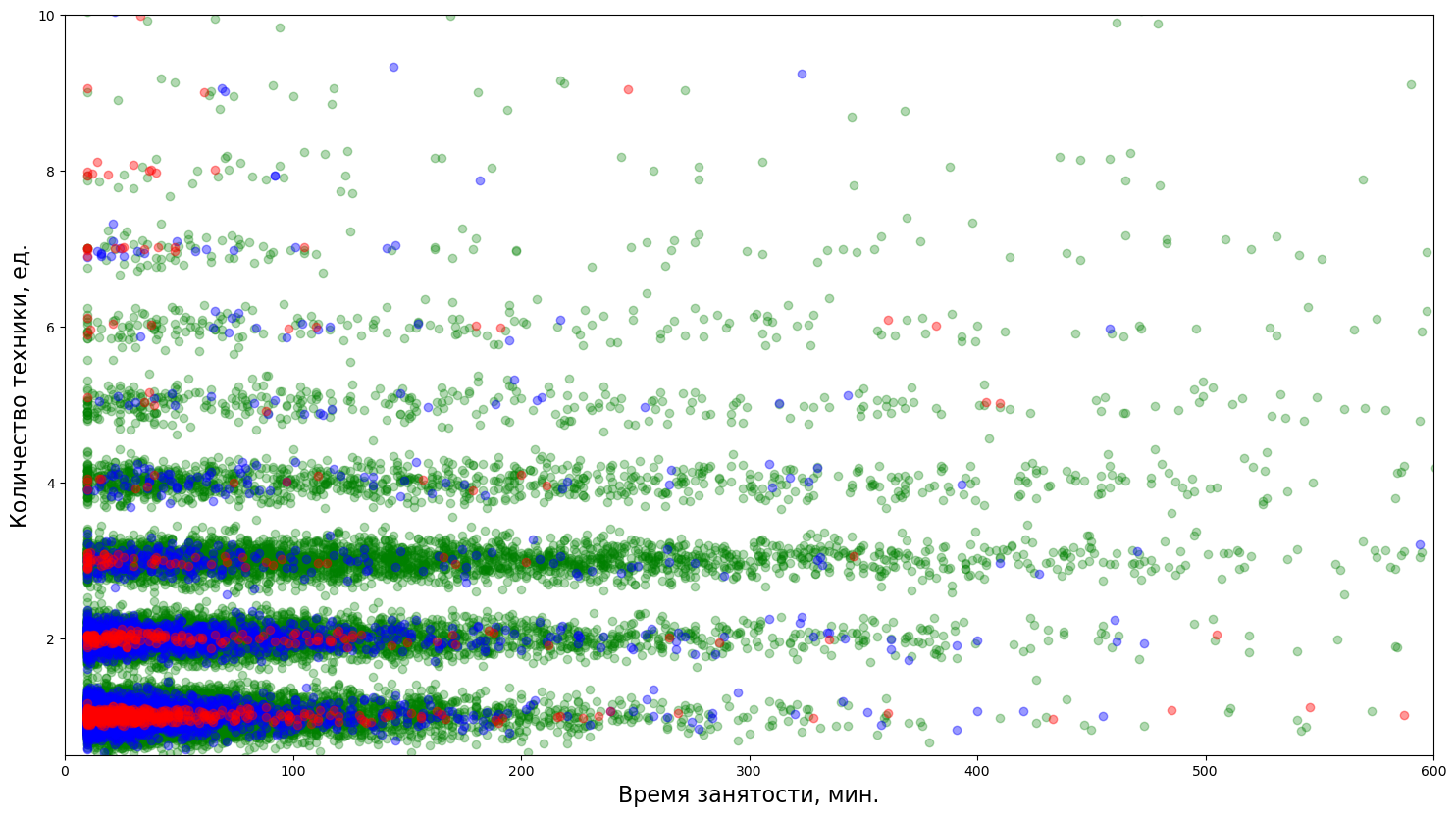


Рисунок 1. Распределение пожаров, произошедших в 2021 году в Магаданской (красные точки), Московской (зеленые точки) и Забайкальском крае (синие точки) по показателям времени занятости (мин.) и количеству привлекаемой техники (ед.). Для наглядности для ординат точек добавлен ε – Гауссовский шум

Впрочем, некоторые закономерности прослеживаются – маловероятно, что в Магаданской области за год произойдет более 20 тыс. пожаров, а для Московской области это практически достоверное событие. Еще одно достоверное событие – во всех регионах Российской Федерации подавляющее большинство пожаров ликвидируются за время, не превышающее одного часа (Рисунок 2). В случае превышения времени, можно предположить, что пожар представляет повышенную тактическую сложность для пожарно-спасательных подразделений.

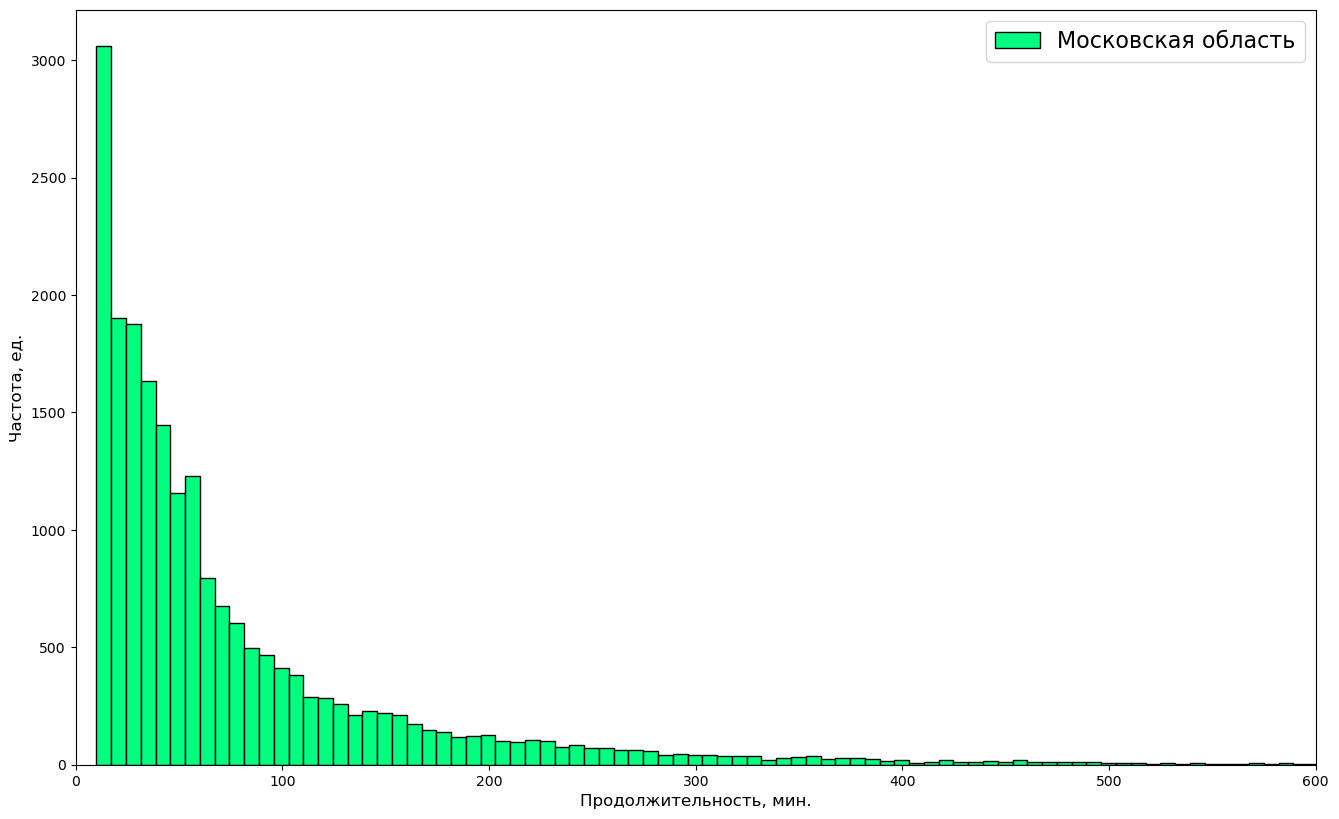


Рисунок 2. Частота распределения пожаров, произошедших в 2021 году в Московской по показателям времени занятости (мин.)

Вместе с тем, показатель длительности тушения пожаров не всегда является единственным подтверждением тактической сложности. Например, горение (тление) торфа или отложений горючих материалов в труднодоступных для подачи огнетушащих веществ местах может длиться сутками, но при этом не возникает необходимости привлечения большого количества сил и средств противопожарной службы и не является сложным.

В ситуациях, когда для успешной ликвидации горения требуется высокая интенсивность действий участников тушения пожара с использованием специальной техники и оборудования в условиях воздействия опасных факторов пожара в течение длительного времени, то такие пожары относятся к затяжным.

Длительность затяжных пожаров принято считать от двух часов и более. Значение определено исходя из того, что время защитного действия современных дыхательных аппаратов, с учетом запасных баллонов, вывозимых на пожарном автомобиле в общей сложности составляет порядка двух часов [1].

Гипотеза исследования, описываемого в статье заключается в подтверждении возможности прогнозирования методами машинного обучения (Machine Learning) оценки возможности перехода ординарного пожара в стадию затяжного.

Под машинным обучением (далее – ML) здесь понимается процесс поиска математической формулы, применяя которую к некоторому набору входных (или обучающих) данных *N* будет получен желаемый результат, например, прогноз или классификация [2].

Для практической реализации гипотезы была сформирована выборка *N* по трем регионам Российской Федерации из 26389 наблюдений[[1]](#footnote-1) и 17 показателей. Распределение количества наблюдений (образцов) по отдельным регионам приведено в начале статьи. Показатели отбирались по принципу формирования некоторого набора первичных признаков для прогнозирования вариантов развития пожара. Подробное описание процесса подготовки данных, создания модели, проектирования признаков и разработки программного кода на языке Python приводится в файле Notebook Jupyter в репозитории - https://github.com/VistaSV30/Long.git.

Алгоритм ML «обучение с учителем» предполагает организацию набора данных в виде коллекции размеченных образцов (1).

|  |  |
| --- | --- |
|  | (1) |

Каждый элемент *xi* из *N* представляет собой вектор признаков отдельного пожара (образца), в котором j = 1, …, D содержатся характеристики единичного признака. Например, x(0) – код региона Российской Федерации (на рисунке 3 это – F1), x(1) – тип населенного пункта, где произошел пожар (F6) и т.д. Метка *yi* является элементом конечного множества классов {1, 2, …, C}, в нашем случае это признак ординарного и затяжного пожара {0,1}. Цель создаваемого алгоритма на основе вектора признаков *x* определить метку *y*.

По корреляционной матрице (Рисунок 3) определяется наличие связей между каждой парой. Сила связей приведена в процентах. Признаки F16 (Степень огнестойкости) и F12 (Объект пожара) имеют сильную отрицательную корреляционную связь, это объясняется сходством признаков. Возможно, допустить, что степень огнестойкости во многом зависит от вида объекта. Это означает, что для прогнозирования длительности пожара один из показателей является избыточным, и может быть удален из выборки.



Рисунок 3. 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

После удаления F16 и других сильно коррелирующих признаков (F15, F15, Exiting, Flow) матрица будет выглядеть следующим образом (Рисунок 4).

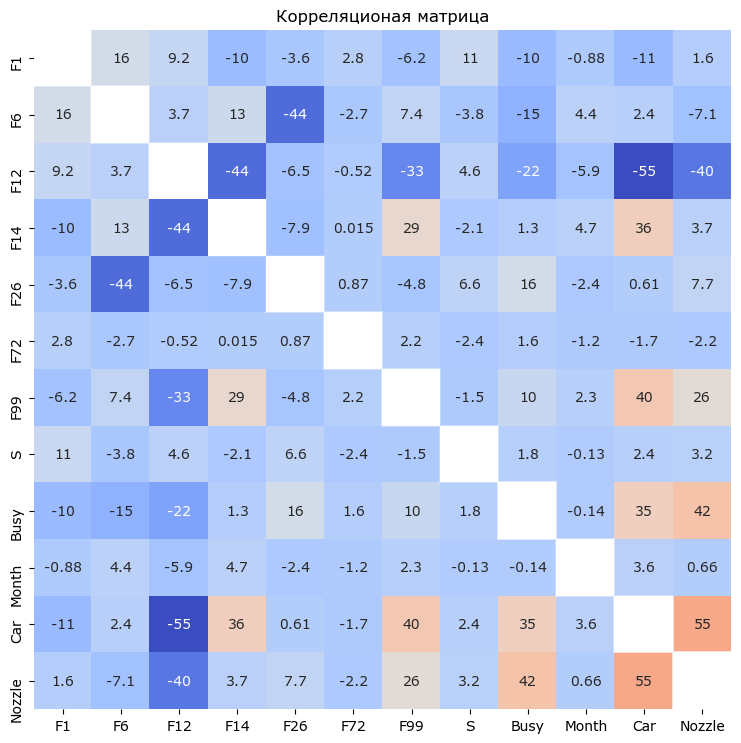


Рисунок 4. Корреляционная матрица после удаления «избыточных» признаков

Оценка выборки на наличие выбросов (даже без Busy и S, явно имеющих очень большие разбросы) показывает, что оставшиеся признаки так же недостаточно однородны (Рисунок 5). Результаты расчетов методом линейной регрессии зависят от неоднородностей признаков, поэтому целесообразно выполнить их нормализацию. Главным условием правильной нормализации, является обеспечение для всех признаков равных возможностей влияния, при этом алгоритмы ML работают лучше и быстрее сходятся [3].

Единого универсального метода нормализации пока не существует. Чаще всего используют один из трех: «Минимакс», «Z-масштабирование» или «Робастная нормализация». В модуле scikit-learn указанные методы соответственно называются MinMaxScaler, StandardScaler и RobustScaler. На основании сравнительного анализа модифицированных распределений (Рисунок 6) и рекомендаций [4] в данном случае принято решение использовать метод StandardScaler.

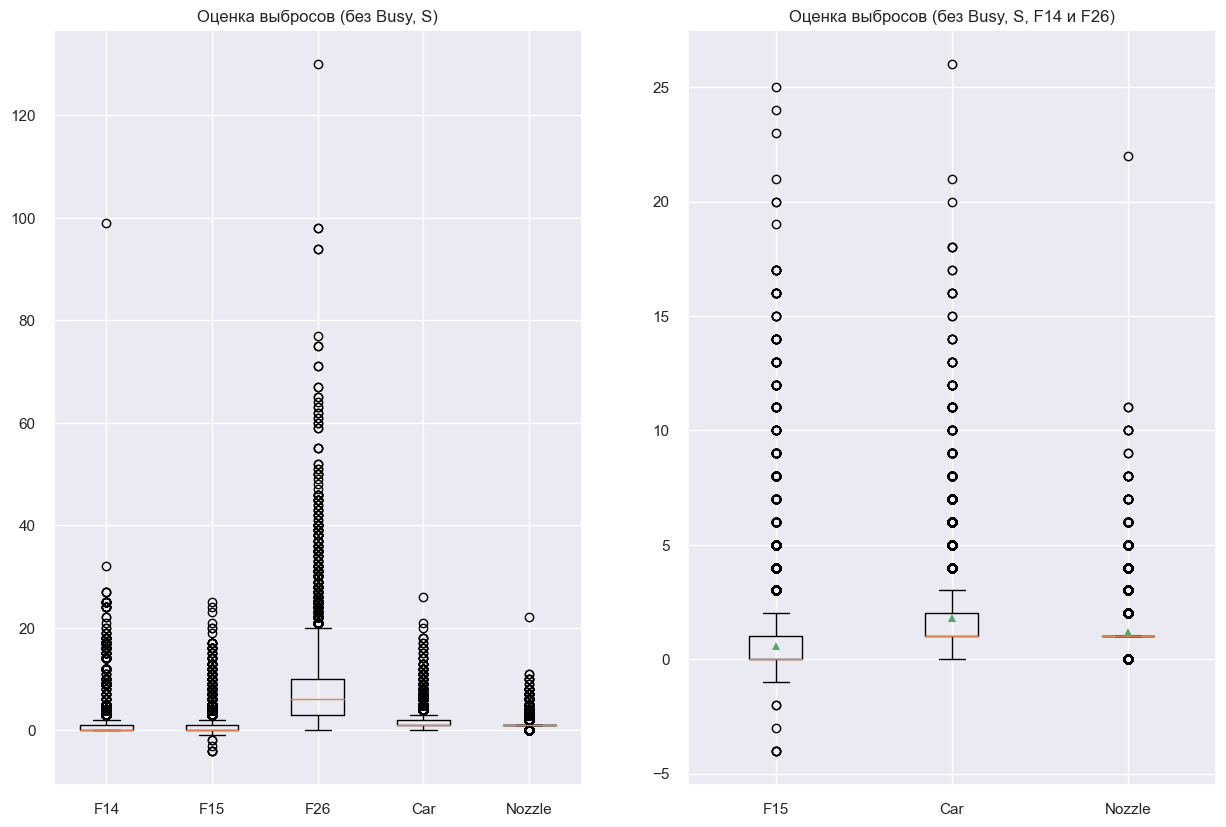


Рисунок 5. Оценка выбросов количественных признаков. Оранжевая полоска – медиана признака. Вершина треугольника – среднее арифметическое

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| Диапазон значений распределения признаков | | | |

Рисунок 6. Варианты нормализации данных методами: MinMaxScaler, RobustScaler и StandardScaler

Распределение числовых признаков после нормализации представлено на диаграмме (Рисунок 7). Теперь даже признаки S и Busy имеют сравнимый с остальными масштаб разброса.



Рисунок 7. Оценка выбросов количественных признаков после нормализации методом StandardScaler

Еще одним важным требованием для корректной работы модели является преобразование категориальных признаков исходной выборки *N* в числовые. Например, признак F6 (Вид населенного пункта) из 8-ми видов используемых в Федеральной базе данных «Пожары» (далее – ФБД), целесообразно свернуть до двух: 1 – города; 0 – сельская местность. Виды объектов пожара (показатель F12) в ФБД имеет 247 категорий, которые основываясь на особенности тактики тушения пожаров и практического опыта можно сгруппировать в пять категорий (Рисунок 7а).

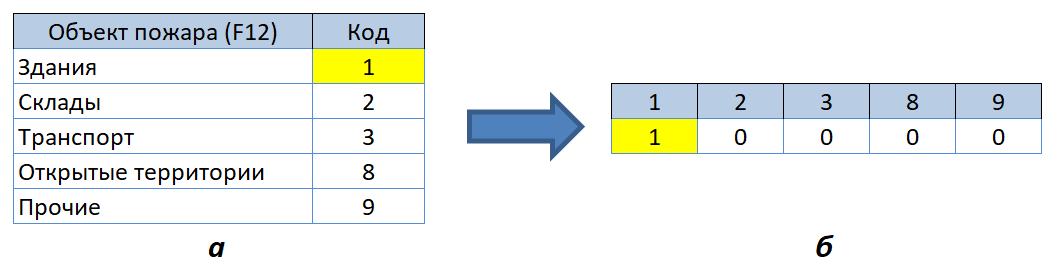


Рисунок 7. Схема преобразования категориального (а) признака в числовой (б)

Это можно пояснить на следующем примере. В исходной выборке *N* для некоторого образца 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. (Рисунок 7б). В результате каждому объекту будет соответствовать значение единицы в соответствующем признаке. В Python такое преобразование выполняется командой Pandas.get\_dummies() с параметром drop\_first= False.

Модель логистической регрессии *fw,b*(*x*) конструируется на основе набора (1) и является функцией от *xi* для бинарных значений *y* (2).

|  |  |
| --- | --- |
|  | (2) |

где w – D-мерный вектор параметров; b – действительное число.

Модель использует для предсказания неизвестного значения метки *y* для данного *x*: . Процесс обучения модели, по сути это поиск оптимальных значений , при которых модель дает наиболее точные прогнозы. Достигается при максимизации критерия правдоподобия (3).

|  |  |
| --- | --- |
|  | (3) |

Для практических расчетов удобнее максимизировать логарифм правдоподобия (4), поскольку ln – строго возрастающая функция максимизация, которой равносильна максимизации ее аргумента.

|  |  |
| --- | --- |
|  | (4) |

Данная задача оптимизации не имеет аналитического решения и поэтому используется процедура численной оптимизации, называемая – градиентный спуск.

Для получения адекватных результатов процесс обучения и тестирования нужно проводить на разных наборах данных. В противном случае невозможно узнать, насколько хорошо модель идентифицирует данные, которые она не получала раньше. С использованием методов библиотеки Scikit-learn, разбиваем исходную выборку (N = 26389) на две части размером 20% и 80% (соответственно 5278 и 21111 образец). Часть в 20% данных предназначена для тестирования (чтобы удалить элемент случайности, при сравнении разных моделей для функции train\_test\_split() будет использован параметр random\_ state=42.

Проверка качества созданной модели логистической регрессии (2) с использованием критерия правдоподобия (4) может быть выполнена несколькими способами. Например, с помощью матрицы запутанности (так же используются названия «диаграмма истинности», «Confusion Matrix») [5].

|  |  |  |  |
| --- | --- | --- | --- |
| Матрица запутанности (Confusion Matrix) | | | |
|  |  | Прогноз | |
|  |  | Ординарный | Затяжной |
| Факт | Ординарный | 5226 (99,0 %) | 9 (0,2 %) |
| Затяжной | 25 (0,5 %) | 18 (0,3 %) |

В 99 % случаев модель ML по тестовой выборке верно определила ординарные пожары. 25 образцов ложно определены как ординарные. Получается, что более половины (25 из 43 затяжных пожаров) распознаны неверно, в репозитории Hithab приведена выборка с этими образцами. Выявленные недостатки возможно устранить методами ML.

На диаграмме (Рисунок 9) показаны признаки и их степень влияния на условия классификации образцов. Наиболее значимые факторы оказывающие влияние на выбор в пользу затяжных пожаров – F99 и Busy. Факторы в пользу ординарных пожаров – F12\_8 и F72\_3.

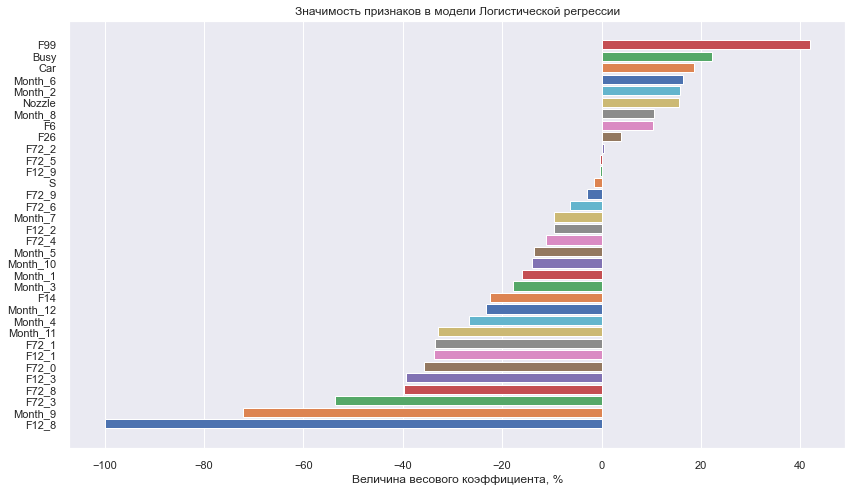


Рисунок 9. Значимость признаков: F99 – использование СИЗОД; Busy – время занятости на пожаре, мин.; Car – общее количество пожарной техники, ед.; Nozzle – общее количество стволов, ед.; F6 – вид населенного пункта; F26 – расстояние до места пожара, км; S – площадь пожара, м2; F14 – этажность здания; 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. Даже в самых «спокойных» регионах, как Магаданская область за год происходит порядка 1000 пожаров. Затяжные пожары представляют интерес для изучения в пожарных подразделениях, а также в научных организациях МЧС России, целях получения положительного опыта и совершенствования деятельности по тушению пожаров. Применение методов машинного обучения позволяет на основе большого объема данных (Big Data) оперативно выявлять факторы и взаимосвязи, оказывающие значимое влияние на процесс пожаротушения и общую организацию деятельности пожарно-спасательных подразделений.

1. 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.
2. Бурков Андрей Машинное обучение без лишних слов. — СПб.: Питер, 2020. — 192 с.: ил. — (Серия «Библиотека программиста»).
3. Харрисон, Мэтт. Машинное обучение: карманный справочник. Краткое руководство по методам структурированного машинного обучения на Python. : Пер. с англ. - СПб. : ООО "Диалектика", 2020 - 320 с. : ил. - Парал. тит. англ.
4. Нолис Ж., Робинсон Э. DataScience для карьериста. – СПб.: Питер, 2021. – 368 с.: ил. – (Серия «Библиотека программиста»).
5. Спринглер В. Гид по Computer Science, расширенное издание. – СПб.: Питер, 2021. – 304 с.: ил. – (Серия «Библиотека программиста»).

1. В ML обычно для наблюдений используется термин «образец», а для показателей – «признак». [↑](#footnote-ref-1)