

Научная статья

УДК 614.8; DOI: 10.61260/2307-7476-2024-4-14-21

ПРОГНОЗИРОВАНИЕ ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЙ И ПРОИСШЕСТВИЙ НА ТЕРРИТОРИИ ПЕРМСКОГО КРАЯ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

✉ Луценко Сергей Георгиевич;

Ташкинов Даниил Сергеевич;

Плещев Виталий Алексеевич.

Главное управление МЧС России по Пермскому краю, г. Пермь, Россия.

Тюрин Герман Владимирович.

Национальный центр управления в кризисных ситуациях МЧС России, Москва, Россия

✉ lucenkosg@gmail.com

Аннотация. В современном мире, где количество данных растет экспоненциально, методы машинного обучения становятся ключевым инструментом для прогнозирования чрезвычайных ситуаций и происшествий. В последнее время ансамблевые алгоритмы активно применяются для решения подобных задач. Одним из наиболее эффективных подходов является градиентный бустинг над решающими деревьями, который сочетает в себе гибкость деревьев и мощь градиентной оптимизации. В данной статье представлена оценка возможностей применения метода градиентного бустинга над решающими деревьями на примере чрезвычайных ситуаций и происшествий на территории Пермского края. Представлены основные преимущества и ограничения применения метода. Проведенные исследования доказали эффективность применения метода градиентного бустинга над решающими деревьями для задач прогнозирования чрезвычайных ситуаций. Его способность учитывать сложные зависимости и работать с разнородными данными, делает его мощным инструментом в арсенале аналитиков и специалистов по прогнозированию.

Ключевые слова: машинное обучение, моделирование, прогнозирование, оценка качества принятия решений

Для цитирования: Луценко С.Г., Ташкинов Д.С., Плещев В.А., Тюрин Г.В. Прогнозирование чрезвычайных ситуаций и происшествий на территории Пермского края с использованием методов машинного обучения // Природные и техногенные риски (физико-математические и прикладные аспекты). 2024. № 4 (52). С. 14–21. DOI: 10.61260/2307-7476-2024-4-14-21.

Scientific article

FORECASTING OF EMERGENCIES AND INCIDENTS ON THE TERRITORY OF THE PERM REGION USING MACHINE LEARNING METHODS

✉ Lutsenko Sergey G.;

Tashkinov Daniil S.;

Pleshcheyev Vitaly A.

Main directorate of EMERCOM of Russia in the Perm region, Perm, Russia.

Tyurin German V.

National crisis management center of EMERCOM of Russia, Moscow, Russia

✉ lucenkosg@gmail.com

Abstract. In the modern world, where the amount of data grows exponentially, machine learning methods are becoming a key tool for predicting emergencies and incidents. Recently, ensemble algorithms have been actively used to solve such problems. One of the most effective approaches is gradient boosting over decision trees, which combines the flexibility of trees and the power of gradient optimization. This article presents an assessment of the possibilities of using the gradient boosting method over decision trees using the example of emergencies and incidents in the Perm Territory. The main advantages and limitations of the method are presented. The conducted studies have proven the effectiveness of the gradient boosting method over decision trees for emergency forecasting tasks. Its ability to take into account complex dependencies and work with heterogeneous data makes it a powerful tool in the arsenal of analysts and forecasting specialists.

Keywords: machine learning, modeling, forecasting, quality assessment of decision-making

For citation: Lutsenko S.G., Tashkinov D.S., Pleshcheyev V.A., Tyurin, G.V. Forecasting of emergencies and incidents on the territory of the Perm region using machine learning methods // Prirodnye i tekhnogennye riski (fiziko-matematicheskie i prikladnye aspekty) = Natural and man-made risks (physico-mathematical and applied aspects). 2024. № 4 (52). P. 14–21. DOI: 10.61260/2307-7476-2024-4-14-21.

Введение

Чрезвычайные ситуации (ЧС) и происшествия, такие как природные катастрофы, аварии на производстве, транспортные инциденты и террористические акты, представляют собой серьезную угрозу для безопасности общества и экономики. Традиционные методы прогнозирования, основанные на статистических моделях, зачастую оказываются недостаточно точными из-за сложности и многомерности данных [1, 2]. В этой связи методы машинного обучения, способные анализировать большие объемы информации и выявлять скрытые закономерности, становятся все более востребованными [3, 4].

Важным аспектом применения методов машинного обучения для прогнозирования является адаптивность машинного обучения. Алгоритмы могут непрерывно обучаться на новых данных, что делает их прогнозы более точными с течением времени. Это особенно полезно в условиях изменяющейся окружающей среды, где традиционные статистические методы могут устаревать [5, 6].

Машинное обучение способно интегрировать разнородные данные, что значительно расширяет возможности анализа. Например, для прогнозирования дорожно-транспортных происшествий (ДТП) можно использовать данные о погоде, трафике, состоянии дорожного покрытия и даже поведении водителей. Такая комплексность позволяет создавать более детализированные и точные прогнозы, что особенно важно для оперативного реагирования и планирования [7].

В рамках федерального проекта «Цифровой прорыв. Сезон: Искусственный интеллект», организованного Министерством экономического развития Российской Федерации и АНО «Россия – страна возможностей» 27–29 октября 2023 г. в г. Перми, прошел окружной хакатон по искусственному интеллекту. Главное управление МЧС России по Пермскому краю приняло участие в качестве кейсодержателя (постановщика задачи).

На основе предоставленного Главным управлением МЧС России по Пермскому краю dataset о происшествиях, архивных данных о погодных условиях и прогнозов погоды на 10 дней участникам было предложено разработать решения по прогнозированию рисков на территории Пермского края.

14 команд участников выполнили кейсовое задание «Прогнозирование опасностей и рисков Пермского края с применением методов машинного обучения». Из 14 команд 11 выполнили задание с использованием градиентного бустинга, один – нейросетью, но с низкими результатами точности прогнозных моделей.

В представленной статье рассматривается оценка возможностей применения методов машинного обучения при прогнозировании рисков ЧС и происшествий в Пермском крае с использованием для обучения моделей Catboost от Яндекс – градиентный бустинг над решающими деревьями.

Цель данного исследования состоит в оценке возможностей применения методов машинного обучения при прогнозировании рисков ЧС и происшествий на примере Пермского края.

Новизна данного исследования заключается в применении современных алгоритмов машинного обучения, таких как градиентный бустинг по решающим деревьям с использованием метеорологических данных и метеопрогнозов для прогнозирования рисков возникновения ЧС и происшествий на территории Пермского края.

Практическая значимость исследования определяется его потенциалом для повышения точности прогнозов и, как следствие, для снижения рисков для населения и экономики, а также для оптимизации ресурсов органов, задействованных в предупреждении и ликвидации ЧС.

Методы исследования

Всего на территории Пермского края с 2012 по 2022 г. произошло 53 137 ЧС и происшествий. Проанализировав распределение по категориям происшествий, было решено уменьшить их количество с 61 до 4 категорий (табл.).

Таблица

Основные категории происшествий на территории Пермского края с 2012 по 2022 г.

№	Тип	Кол-во
1	ДТП	25 538
2	Взрывы/пожары/разрушения	25 156
3	ЖКХ	1 790
4	Аварии на энергетических сетях	210

На основе этих данных был подготовлен набор данных (dataset). Для обучения моделей использовался Catboost от Яндекс – градиентный бустинг над решающими деревьями.

CatBoost – это метод машинного обучения, разработанный компанией Яндекс, который основан на градиентном бустинге над решающими деревьями. Он отличается высокой эффективностью и точностью, особенно при работе с категориальными данными. Одной из ключевых особенностей CatBoost является автоматическая обработка категориальных признаков, что избавляет пользователя от необходимости предварительного кодирования, такого как one-hot encoding или label encoding. Это достигается за счет использования оригинального подхода, основанного на перестановках и статистиках, что позволяет модели эффективно учитывать взаимодействия между категориальными и числовыми признаками [8].

Важным преимуществом CatBoost является его устойчивость к переобучению. Это достигается благодаря использованию упорядоченного бустинга, который учитывает порядок объектов в данных для вычисления градиентов. Такой подход позволяет минимизировать смещение, возникающее при использовании традиционного градиентного бустинга. Кроме того, CatBoost поддерживает различные функции потерь, что делает его универсальным инструментом для решения задач классификации, регрессии и ранжирования [9].

CatBoost также предлагает встроенные механизмы для работы с пропущенными значениями и автоматической настройки гиперпараметров [10]. Это значительно упрощает процесс разработки моделей, особенно для пользователей, которые не обладают глубокими знаниями в области машинного обучения. Модель поддерживает многопоточность и может эффективно работать на больших объемах данных, что делает её применимой как для небольших проектов, так и для крупномасштабных задач.

Ещё одной отличительной чертой CatBoost является его интерпретируемость. Модель предоставляет инструменты для анализа важности признаков, что позволяет лучше понять, какие факторы влияют на предсказания.

Для исключения случаев переобучения модели в исследовании использовалась кросс-валидация – это метод оценки производительности модели, который помогает уменьшить вероятность переобучения. Он включает разделение данных на несколько подмножеств (или фолдов) и последующее обучение и оценку модели несколько раз.

Стратифицированная кросс-валидация, которая гарантирует, что каждое разбиение данных будет иметь ту же пропорцию классов, что и полный набор данных.

Также была построена матрица ошибок (confusion matrix) – это инструмент, используемый в машинном обучении для оценки производительности классификационной модели. Она показывает, какие решения принимает модель, сравнивая их с истинными значениями. Матрица ошибок состоит из четырех частей:

1. Истинно положительные (True Positives, TP): количество случаев, когда модель правильно предсказала положительный класс.

2. Истинно отрицательные (True Negatives, TN): количество случаев, когда модель правильно предсказала отрицательный класс.

3. Ложно положительные (False Positives, FP): количество случаев, когда модель неправильно предсказала положительный класс (ошибки первого рода).

4. Ложно отрицательные (False Negatives, FN): количество случаев, когда модель неправильно предсказала отрицательный класс (ошибки второго рода).

Интерпретация матрицы ошибок включает в себя следующие аспекты:

– Точность (Accuracy): общая доля правильных предсказаний из всех. Вычисляется как $(TP + TN) / (TP + TN + FP + FN)$.

– Чувствительность (Sensitivity) или Полнота (Recall): доля правильно идентифицированных положительных случаев из всех истинных положительных. Вычисляется как $TP / (TP + FN)$.

– Специфичность (Specificity): доля правильно идентифицированных отрицательных случаев из всех истинных отрицательных. Вычисляется как $TN / (TN + FP)$.

– Точность (Precision): доля истинно положительных предсказаний из всех положительных предсказаний. Вычисляется как $TP / (TP + FP)$.

– F1-мера: гармоническое среднее между точностью и полнотой. Вычисляется как $2 * (Precision * Recall) / (Precision + Recall)$.

Интерпретация матрицы ошибок позволяет понять, насколько хорошо модель работает в различных аспектах, включая ее способность правильно идентифицировать каждый класс и ее общую точность. Это также помогает выявить, может ли модель быть смещенной в сторону одного из классов (например, предсказывать слишком много положительных или отрицательных результатов).

Результаты исследования и их обсуждение

Результат обучения на обучающей выборке в рамках работы получился равным 0,7506476.

Итого показатель «Средняя F1-score» по типам происшествий:

– опасные природные явления – 0.6211067;

– прочие опасности – 0.7089366;

– ЖКХ – 0.7344406;

– нет события – 0.78629965;

– взрывы/пожары/разрушения – 0.7531637;

– ДТП – 0.8049544 (рис. 1);

– аварии с выбросом опасных/токсичных веществ – 0.8456318.

Модели в целом хорошо предсказывают соответствующего типа события.

Соответственно, этот метод подходит для прогнозирования ЧС и происшествий.

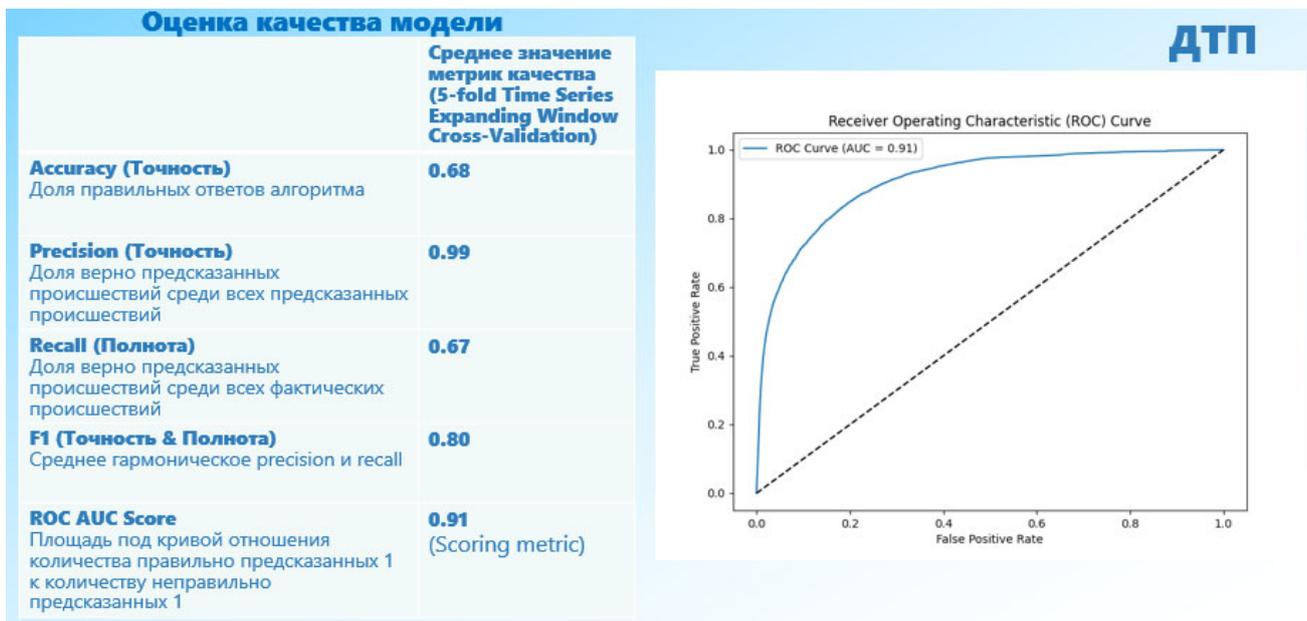


Рис. 1. Оценка качества модели прогнозирования ДТП

В работе была использована библиотека SHAP (SHapley Additive exPlanations), которая помогает понять вклад каждого признака в предсказания модели. Визуализируются SHAP-значения, чтобы показать, как признаки влияют на модель в целом. График суммарных SHAP-значений (summary_plot) позволяет увидеть распределение влияния признаков на все предсказания модели.

SHAP-значения являются мощным инструментом для интерпретации сложных моделей машинного обучения. Они помогают объяснить, почему конкретные прогнозы были сделаны, выявляя вклад каждого признака в итоговое предсказание модели. На графике отображаются признаки в порядке убывания их важности, каждый признак представлен горизонтальной линией, длина которой соответствует важности признака. Цвет точек показывает значение признака (высокое или низкое).

Объяснение предсказаний: SHAP-значения показывают, насколько каждый признак увеличивает или уменьшает предсказанное значение по сравнению со средним. Они могут быть положительными (увеличивают вероятность исхода) или отрицательными (уменьшают её).

По результатам анализа ключевыми факторами для типа ДТП являются (рис. 2):

1. Видимость.
2. Влажность и температура.
3. Накопленное количество холодных дней.
4. Накопленное количество дней без осадков.
5. Признак сезонной цикличности.
6. Давление.



Рис. 2. Результаты анализа важности факторов, влияющих на динамику ДТП

В процессе обучения моделей искусственного интеллекта были использованы данные за период с 2012 по 2023 г.: около 300 млн метеонаблюдений с учетом метеопостов Пермского края, 52 тыс. инцидентов по четырем группам (ДТП, техногенные пожары, аварии на коммунальных сетях водоснабжения и водоотведения, аварии на электросетях).

Заключение

Методы машинного обучения открывают новые возможности для прогнозирования ЧС и происшествий. Они позволяют не только повысить точность прогнозов, но и снизить затраты на предотвращение и ликвидацию последствий.

В результате апробации на данных Пермского края удалось получить высокоточные модели взаимосвязи природных факторов (точность моделей составляет от 85 % до 94 %).

В работе были достигнуты все поставленные задачи по прогнозированию опасностей и рисков в соответствии с данными о прогнозируемой погоде.

Реализовано применение метода машинного обучения в модели прогнозирования ЧС и происшествий. Полученные результаты позволяют уверенно определять возможные риски на территории Пермского края.

Прототип разработанной системы с использованием результатов разработанной модели в качестве пилотного проекта будет рекомендован к прохождению опытной эксплуатации в центре управления кризисных ситуаций Главного управления МЧС России по Пермскому краю в 2024–2025 гг.

Одним из ключевых направлений дальнейших исследований является интеграция методов машинного обучения с системами реального времени. Это позволит не только прогнозировать ЧС, но и оперативно реагировать на них. Например, разработка адаптивных моделей, способных учитывать изменения в данных в режиме реального времени, может значительно повысить эффективность систем мониторинга и предотвращения аварий.

Применение методов машинного обучения способствует автоматизации процессов анализа и принятия решений. Это не только снижает нагрузку на специалистов, но и позволяет быстрее реагировать на потенциальные угрозы. Например, системы на основе искусственного интеллекта могут автоматически генерировать предупреждения и рекомендации для служб спасения, что значительно ускоряет процесс реагирования [11]. Таким образом, машинное обучение становится неотъемлемым инструментом для повышения безопасности и минимизации рисков ЧС.

Список источников

1. Матвеев А.В. Методы моделирования и прогнозирования. СПб.: С.-Петербург. ун-т ГПС МЧС России, 2022. 230 с. ISBN 978-5-907116-73-3. EDN IMLKWS.
2. Матвеев А.В., Богданова Е.М. Классификация методов прогнозирования чрезвычайных ситуаций // Национальная безопасность и стратегическое планирование. 2018. № 4 (24). С. 61–70. EDN YTPZYL.
3. Сапожников А.А. Искусственный интеллект на службе МЧС России // Гражданская защита. 2023. № 5 (573). С. 28–29.
4. Луценко С.Г., Уткин О.В. Цифровые технологии Единой государственной системы предупреждения и ликвидации чрезвычайных ситуаций: анализ решений и перспективы развития // Природные и техногенные риски (физико-математические и прикладные аспекты). 2023. № 3 (47). С. 22–28. DOI: 10.61260/2307-7476-2023-3-22-28. EDN VMMDNC.
5. Machine learning in disaster management: recent developments in methods and applications / Linardos V. [et al.] // Machine Learning and Knowledge Extraction. 2022. Vol. 4. № 2. DOI: 10.3390/make4020020.
6. Ключева И.А. Современные возможности и примеры внедрения машинного обучения // Оригинальные исследования. 2021. Т. 11. № 7. С. 12–32. EDN HBQRYC.
7. Huang D., Wang S., Liu Z. A systematic review of prediction methods for emergency management // International Journal of Disaster Risk Reduction. 2021. Vol. 62. P. 102412. DOI: 10.1016/j.ijdr.2021.102412.
8. Hancock J.T., Khoshgoftaar T.M. CatBoost for big data: an interdisciplinary review // Journal of Big Data. 2020. Vol. 7. № 1. P. 94. DOI: 10.1186/s40537-020-00369-8.
9. CatBoost model and artificial intelligence techniques for corporate failure prediction / S.B. Jabeur [et al.] // Technological Forecasting and Social Change. 2021. Vol. 166. P. 120658. DOI: 10.1016/j.techfore.2021.120658.
10. Comparison of the CatBoost Classifier with other Machine Learning Methods / A.A. Ibrahim [et al.] // International Journal of Advanced Computer Science and Applications. 2020. Vol. 11. № 11. P. 738–748. DOI: 10.14569/IJACSA.2020.0111190.
11. Бородушко И.В., Матвеев А.В. Современные тенденции и стратегические цели развития искусственного интеллекта в Российской Федерации // Национальная безопасность и стратегическое планирование. 2024. № 2 (46). С. 66–74. DOI: 10.37468/2307-1400-2024-2-66-74. EDN EFWHYT.

References

1. Matveev A.V. Metody modelirovaniya i prognozirovaniya. SPb.: S.-Peterb. un-t GPS MCHS Rossii, 2022. 230 s. ISBN 978-5-907116-73-3. EDN IMLKWS.
2. Matveev A.V., Bogdanova E.M. Klassifikaciya metodov prognozirovaniya chrezvychajnyh situacij // Nacional'naya bezopasnost' i strategicheskoe planirovanie. 2018. № 4 (24). S. 61–70. EDN YTPZYL.
3. Sapozhnikov A.A. Iskusstvennyj intellekt na sluzhbe MCHS Rossii // Grazhdanskaya zashchita. 2023. № 5 (573). S. 28–29.
4. Lucenko S.G., Utkin O.V. Cifrovye tekhnologii Edinoj gosudarstvennoj sistemy preduprezhdeniya i likvidacii chrezvychajnyh situacij: analiz reshenij i perspektivy razvitiya // Prirodnye i tekhnogennye riski (fiziko-matematicheskie i prikladnye aspekty). 2023. № 3 (47). S. 22–28. DOI: 10.61260/2307-7476-2023-3-22-28. EDN VMMDNC.
5. Machine learning in disaster management: recent developments in methods and applications / Linardos V. [et al.] // Machine Learning and Knowledge Extraction. 2022. Vol. 4. № 2. DOI: 10.3390/make4020020.
6. Klyueva I.A. Sovremennye vozmozhnosti i primery vnedreniya mashinnogo obucheniya // Original'nye issledovaniya. 2021. T. 11. № 7. S. 12–32. EDN HBQRYC.
7. Huang D., Wang S., Liu Z. A systematic review of prediction methods for emergency management // International Journal of Disaster Risk Reduction. 2021. Vol. 62. P. 102412. DOI: 10.1016/j.ijdr.2021.102412.

8. Hancock J.T., Khoshgoftaar T.M. CatBoost for big data: an interdisciplinary review // Journal of Big Data. 2020. Vol. 7. № 1. P. 94. DOI: 10.1186/s40537-020-00369-8.
9. CatBoost model and artificial intelligence techniques for corporate failure prediction / S.B. Jabeur [et al.] // Technological Forecasting and Social Change. 2021. Vol. 166. P. 120658. DOI: 10.1016/j.techfore.2021.120658.
10. Comparison of the CatBoost Classifier with other Machine Learning Methods / A.A. Ibrahim [et al.] // International Journal of Advanced Computer Science and Applications. 2020. Vol. 11. № 11. P. 738–748. DOI: 10.14569/IJACSA.2020.0111190.
11. Borodushko I.V., Matveev A.V. Sovremennye tendencii i strategicheskie celi razvitiya iskusstvennogo intellekta v Rossijskoj Federacii // Nacional'naya bezopasnost' i strategicheskoe planirovanie. 2024. № 2 (46). S. 66–74. DOI: 10.37468/2307-1400-2024-2-66-74. EDN EFWHYT.

Информация о статье:

Статья поступила в редакцию: 20.10.2024; одобрена после рецензирования: 11.11.2024; принята к публикации: 13.11.2024

Information about the article:

The article was submitted to the editorial office: 20.10.2024; approved after review: 11.11.2024; accepted for publication: 13.11.2024

Информация об авторах:

Луценко Сергей Георгиевич, начальник отдела автоматизации процессов управления центра управления в кризисных ситуациях Главного управления МЧС России по Пермскому краю (614015, г. Пермь, ул. Екатерининская, д. 53а), e-mail: gis.cuks@59.mchs.gov.ru, lucenkosg@gmail.com

Ташкинов Даниил Сергеевич, главный специалист отдела автоматизации процессов управления центра управления в кризисных ситуациях Главного управления МЧС России по Пермскому краю (614015, г. Пермь, ул. Екатерининская, д. 53а), e-mail: tashkinov_ds@59.mchs.gov.ru

Плещев Виталий Алексеевич, главный специалист отдела автоматизации процессов управления центра управления в кризисных ситуациях Главного управления МЧС России по Пермскому краю (614015, г. Пермь, ул. Екатерининская, д. 53а), e-mail: pleshev_va@59.mchs.gov.ru

Тюрин Герман Владимирович, инспектор отдела пространственных данных Управления космического мониторинга Главного управления «Национальный центр управления в кризисных ситуациях» (121357, Москва, ул. Ватутина, д. 1), e-mail: g.tyurin@mchs.gov.ru

Information about the authors:

Lutsenko Sergey G., head of the department of automation of control processes of the crisis management centre of the Main directorate of EMERCOM of Russia for the Perm krai (614015, Perm, Ekaterininskaya str., 53a), e-mail: gis.cuks@59.mchs.gov.ru, lucenkosg@gmail.com

Tashkinov Daniil S., chief specialist of the department of automation of control processes of the crisis management centre of the Main directorate of EMERCOM of Russia for the Perm krai (614015, Perm, Ekaterininskaya str., 53a), e-mail: tashkinov_ds@59.mchs.gov.ru

Pleshchev Vitaly A., chief specialist of the department of automation of control processes of the crisis management centre of the Main directorate of EMERCOM of Russia for the Perm krai (614015, Perm, Ekaterininskaya str., 53a), e-mail: pleshev_va@59.mchs.gov.ru;

Tyurin German V., inspector of the spatial data department of the space monitoring department of the Main Directorate «National Crisis Management Center» (121357, Moscow, Vatutina str., 1), e-mail: g.tyurin@mchs.gov.ru