

УДК: 911.9

DOI: 10.35595/2414-9179-2024-2-30-445-455

Л. А. Израилев<sup>1</sup>, Л. В. Гордиенко<sup>2</sup>

## ГЕОИНФОРМАЦИОННЫЙ МЕТОД ВЫЯВЛЕНИЯ И АНАЛИЗА УЧАСТКОВ КОНЦЕНТРАЦИИ ДОРОЖНО-ТРАНСПОРТНЫХ ПРОИСШЕСТВИЙ

### АННОТАЦИЯ

Согласно данным Всемирной организации здравоохранения ежегодно в автокатастрофах погибает около 1,25 млн чел., а 30–50 млн чел. получают серьезные травмы. В Российской Федерации эта цифра за последние 10 лет в среднем составляла около 17 000 погибших и 201 000 раненых в год. В то же время количество автомобилей на дорогах с каждым годом только увеличивается. В этих условиях задача уменьшения количества дорожно-транспортных происшествий (ДТП) и тяжести их последствий остается актуальной. В данной работе предложен метод определения и анализа участков концентрации дорожно-транспортных происшествий (УКДТП) с учетом пространственной привязки события и исследования территории. Участки с однородными пространственными характеристиками объединяются в кластеры и затем анализируются для того, чтобы лучше понять природу возникновения участков концентрации ДТП. Понимание причин появления участков повышенной аварийности будет способствовать более точному прогнозированию данных участков, а также даст возможность проводить более рациональные мероприятия по уменьшению опасностей на данных участках. В результате это позволит снизить количество транспортных происшествий и их ущерб. В работе применяются различные алгоритмы кластеризации на основе плотности: DBSCAN, HDBSCAN, OPTICS. Каждый из указанных алгоритмов проверяется в двух вариантах значений расстояния поиска и минимального числа объектов на кластер. Для проведения анализа ДТП предлагается использование программного обеспечения геоинформационной системы (ГИС) ArcGIS Pro и данных о местоположении ДТП. ГИС ArcGIS Pro имеет мощные функции пространственного анализа, визуализации, обработки пространственных данных. В статье также делается вывод о том, какой алгоритм кластеризации на основе плотности лучше всего подходит для анализа транспортных происшествий.

**КЛЮЧЕВЫЕ СЛОВА:** дорожно-транспортное происшествие, кластерный анализ, пространственный анализ, ГИС

---

<sup>1</sup> Южный федеральный университет, ул. Шевченко, д. 2, Таганрог, Россия, 347900, *e-mail*: [izrailev@sfedu.ru](mailto:izrailev@sfedu.ru)

<sup>2</sup> Южный федеральный университет, ул. Шевченко, д. 2, Таганрог, Россия, 347900, *e-mail*: [lgordienko@sfedu.ru](mailto:lgordienko@sfedu.ru)

Lev A. Izrailev<sup>1</sup>, Larisa V. Gordienko<sup>2</sup>

## GEOINFORMATION METHOD FOR IDENTIFYING AND ANALYZING AREAS OF CONCENTRATION OF TRAFFIC ACCIDENTS

### ABSTRACT

According to the World Health Organization, about 1.25 million people die each year in car accidents, and 30–50 million people get seriously injured. In the Russian Federation, this figure has averaged about 17 000 deaths and 201 000 injuries per year over the past 10 years. At the same time, the number of cars on the roads is only increasing every year. In these conditions, the task of reducing the number of road traffic accidents and the severity of their consequences remains relevant. This paper proposes a method for identifying and analyzing areas where traffic accidents are concentrated, taking into account the spatial reference of the event and study of the territory. Areas with homogeneous spatial characteristics were combined into clusters and then analyzed in order to better understand the nature of the occurrence of accident concentration areas. Understanding the reasons for the emergence of areas with high accident rates would contribute to more accurate forecasting for these areas. This will also make it possible to carry out more rational measures to reduce hazards in these areas. As a result, this will reduce the number of transport accidents and their damage. The work uses various density-based clustering algorithms: DBSCAN, HDBSCAN, OPTICS. Each of these algorithms is tested in two variants of the search distance and the minimum number of objects per cluster. For accident analysis, it was proposed to use ArcGIS Pro geographic information system (GIS) software and accident location data. GIS ArcGIS Pro has powerful functions for spatial analysis, visualization, and spatial data processing. The paper also concludes which density-based clustering algorithm would be best suited for traffic accident analysis.

**KEYWORDS:** traffic accident, cluster analysis, spatial analysis, GIS

### ВВЕДЕНИЕ

Проблема анализа и снижения дорожно-транспортных происшествий (ДТП) возникла еще в начале XX в., но при этом является актуальной и в настоящее время.

В 1949 г. Рубен Смед [Smeed, 1949] на основе анализа данных по аварийности 20 стран Европы (начиная с 1938 г.) предложил регрессионное уравнение, которое позже легло в основу так называемого «Закона Смида» — эмпирического правила, которое гласило, что количество погибших в ДТП обратно пропорционально количеству собственных легковых автомобилей на 1 000 чел. населения (уровень автомобилизации). Данный закон стал одним из первых способов прогнозирования смертности от ДТП. Впоследствии данный закон был опровергнут в силу того, что во многих странах наблюдалось явление снижения смертности населения в ДТП при увеличении уровня автомобилизации, а также того, что он не учитывал другие факторы, влияющие на смертность в ДТП. Однако в то же время отмечалась [Al-Haji, 2001], что уравнение из закона способно дать результаты, близкие к фактическим.

Для Российской Федерации долгое время оставался характерен более высокий уровень смертности на дорогах, чем прогнозируемый исходя из-за закономерностей, установленных Рубеном Смедом. Развитие дорожной сети не поспевало за ростом числа автомобилей как с точки зрения обеспечения безопасности движения, так и с точки зрения пропускной способности.

<sup>1</sup> Southern Federal University, Shevchenko str., 2, Taganrog, 347900, Russia, e-mail: [izrailev@sfedu.ru](mailto:izrailev@sfedu.ru)

<sup>2</sup> Southern Federal University, Shevchenko str., 2, Taganrog, 347900, Russia, e-mail: [lgordienko@sfedu.ru](mailto:lgordienko@sfedu.ru)

Не все транспортные происшествия происходят в случайных местах. Некоторые ДТП происходят на одних и тех же сегментах дорожной сети, создавая тем самым участки концентрации дорожно-транспортных происшествий (УКДТП), где число аварий с участием автомобилей превышает среднестатистический показатель для аналогичных участков. Помимо участков концентрации с очевидными недостатками эксплуатационного состояния, существуют участки, которые содержат характеристики, косвенно влияющие на нарушение ПДД участниками дорожного движения. Многие из них носят пространственный характер. Например, в условиях недостаточного количества пешеходных переходов или их нерационального расположения возрастает количество случаев перехода дороги пешеходами в неположенных местах, что создает угрозу возникновения ДТП. Своевременное выявление местоположений данных участков, а также формирование гипотез о природе их возникновения — это путь к снижению числа ДТП и повышению безопасности дорожного движения (БДД).

## МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Одним из способов повышения безопасности на дорогах является выявление УКДТП, а также участков повышенной опасности (УПО) на дорогах. К ним относятся участки автомобильной дороги, не превышающие 1 000 м, вне населенного пункта, 200 м — в населенном пункте, или перекрестки дорог, где в течение последних 12 месяцев произошло несколько ДТП по 2 критериям:

- 3 и более ДТП одного вида;
- 5 и более ДТП независимо от их вида, в результате которых погибли или ранены люди<sup>1</sup>.

Выявление УКДТП — это определение местоположения (адрес и (или) географические координаты начала и конца) УКДТП и его характеристик (степени опасности, стабильности, видимости и т. п.)

Поиск участков концентрации ДТП подразумевает статистический и пространственный анализ данных происшествий с целью определения мест наибольшего их распространения. Среди применяемых методов можно выделить метод конфликтных точек, однако чаще всего применяется топографический или математический анализ для выявления наиболее опасных участков дорог [Елисеев и др., 2021].

Для повышения эффективности выявления участков ДТП в данной работе предлагается использование кластерного анализа на основе плотности. Данный метод анализа с помощью алгоритма DBSCAN использовался для выявления УПО на дорогах штата Массачусетс, США в 2013–2018 гг. [Герштейн, Терехов, 2021].

Главное назначение кластерного анализа — разбиение множества исследуемых объектов и признаков на однородные группы или кластеры [Arjun, 2020]. Это означает, что решается задача классификации данных и выявления соответствующей структуры в ней [Haoxiang, 2022]. Методы кластерного анализа можно применять в самых различных случаях, даже в тех случаях, когда речь идет о простой группировке, в которой все сводится к образованию групп по количественному сходству [Nisha, Kaur, 2015]. Большое достоинство кластерного анализа в том, что он позволяет производить разбиение объектов не по одному параметру, а по целому набору признаков. Кроме того, кластерный анализ в отличие от большинства математико-статистических методов не накладывает никаких ограничений на вид рассматриваемых объектов и позволяет рассматривать множество

<sup>1</sup> ОДМ 218.6.015–2015. Рекомендации по учету и анализу дорожно-транспортных происшествий на автомобильных дорогах Российской Федерации. Издан на основании распоряжения Федерального дорожного агентства от 12.05.2015 № 853-р. М., 2015. 78 с.

исходных данных практически произвольной природы. Это имеет большое значение, например, для прогнозирования конъюнктуры, когда показатели имеют разнообразный вид, затрудняющий применение традиционных эконометрических подходов.

Для проведения анализа числа ДТП предлагается использование программного обеспечения ArcGIS Pro и данных о местоположении ДТП. Точность анализа будет напрямую зависеть от точности и полноты исходных сведений [Soloviev et al., 2016].

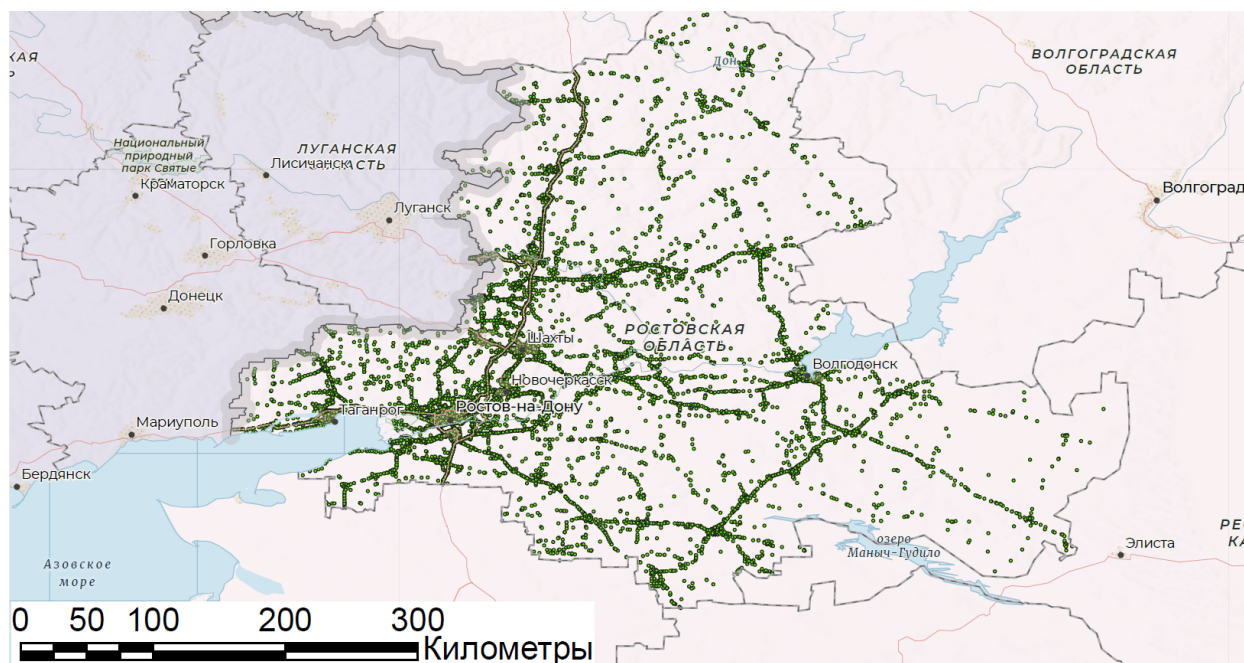
Для анализа данных был использован инструмент «Кластеризация на основе плотности». Принцип работы «Кластеризации на основе плотности» основан на определении области концентрации точек с отделением их от пустых и разреженных областей. Точки, не являющиеся частью кластера, помечаются, как «шум». В качестве метрики расстояния используется евклидова метрика. Для решения задачи по определению УПО в работе были использованы следующие алгоритмы:

- 1) DBSCAN (Density-based spatial clustering of applications with noise) — использует указанное расстояние для отделения плотных кластеров от окружающего шума [Prasad, Srikanth, 2023]. Алгоритм DBSCAN — самый быстрый из рассматриваемых методов кластеризации, но применим только в случае, если известно, какое расстояние поиска необходимо использовать для получения хороших результатов для всех потенциальных кластеров. Это происходит в ситуации сходных плотностей всех важных кластеров.
- 2) HDBSCAN (Hierarchical density-based spatial clustering of applications with noise) — выбирает, на каком уровне кластеров в каждом наборе вложенных кластеров будут созданы наиболее стабильные кластеры, включающие максимально возможное количество объектов и не содержащие шумов. Чем больше будет указанное «минимальное число объектов на кластер», тем более крупные зоны аварийности будут получены. Данный алгоритм самый удобный в использовании за счет минимального количества исходных данных.
- 3) OPTICS (Ordering points to identify the clustering structure) — в данном алгоритме была исправлена существенная слабость алгоритма DBSCAN, а именно проблема идентификации содержательных кластеров в данных, имеющих различные плотности. Используя расстояние между соседними объектами для создания графика достижимости, OPTICS отделяет кластеры различных плотностей от шума. Благодаря этому OPTICS предлагает наибольшую гибкость в тонких настройках определения кластеров, регулируя плотность построенных кластеров, хотя и требует более интенсивных вычислений, особенно с большим расстоянием поиска кластеров.

## РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ И ИХ ОБСУЖДЕНИЕ

В качестве примера были взяты данные о ДТП по Ростовской области с 2015 по 2022 гг. включительно. Результат пространственной привязки данных ДТП приведен на рис. 1. Из-за масштаба дорожная сеть Ростовской области и, в частности, г. Таганрога отражена не в полном объеме. По этой причине не для всех объектов ДТП присутствует обозначение дороги.

Исходные данные включают следующую информацию: координаты, время суток, близость к искусственным объектам, административный район, адрес, погодные условия, категория ДТП, дата, количество погибших и число пострадавших, степень тяжести ранений у пострадавших, информация об автомобиле и установленных нарушениях ПДД, состояние дорожной инфраструктуры, участники ДТП.



*Рис. 1. Исходные данные о ДТП*  
*Fig. 1. Initial data on accidents*

*Этап 1. Применение алгоритмов кластеризации на основе плотности для определения УПО на дорогах*

Для выявления наиболее опасных участков дорожной сети применялись следующие виды алгоритмов кластеризации: DBSCAN, HDBSCAN и OPTICS. Каждый алгоритм был проверен в следующих вариантах:

*Вариант 1.* Поиск при значениях расстояния поиска и минимального числа объектов на кластер 1 000 м (1 км) и 50 объектов, соответственно, в Ростовской области (данный режим позволит определить населенные пункты и их районы с наибольшей концентрацией ДТП).

*Вариант 2.* Поиск при значениях расстояния поиска и минимального числа объектов на кластер 100 м и 10 объектов, соответственно, в г. Таганроге (данный режим соответствует поиску участков концентрации ДТП, однако из-за особенностей работы алгоритма и условий поиска данные значения были изменены для получения результатов, соответствующих нормативам).

Алгоритм DBSCAN при значениях варианта 1 построил кластеры аварийности в следующих городах Ростовской области: Таганрог, Ростов-на-Дону, Батайск, Чалтырь, Азов, Новочеркасск, Новошахтинск, Шахты, Донецк, Каменск-Шахтинский, Волгодонск, Сальск. Результат работы алгоритма DBSCAN согласно варианту 1 приведен на рис. 2.

При значениях Варианта 2 алгоритм построил кластеры аварийности на одних из самых оживленных перекрестках и участках дорожной сети г. Таганрог. За пределами города был построен кластер на пересечении трасс А-280 и 60к-232. Результат работы алгоритма DBSCAN согласно варианту 2 приведен на рис. 3.

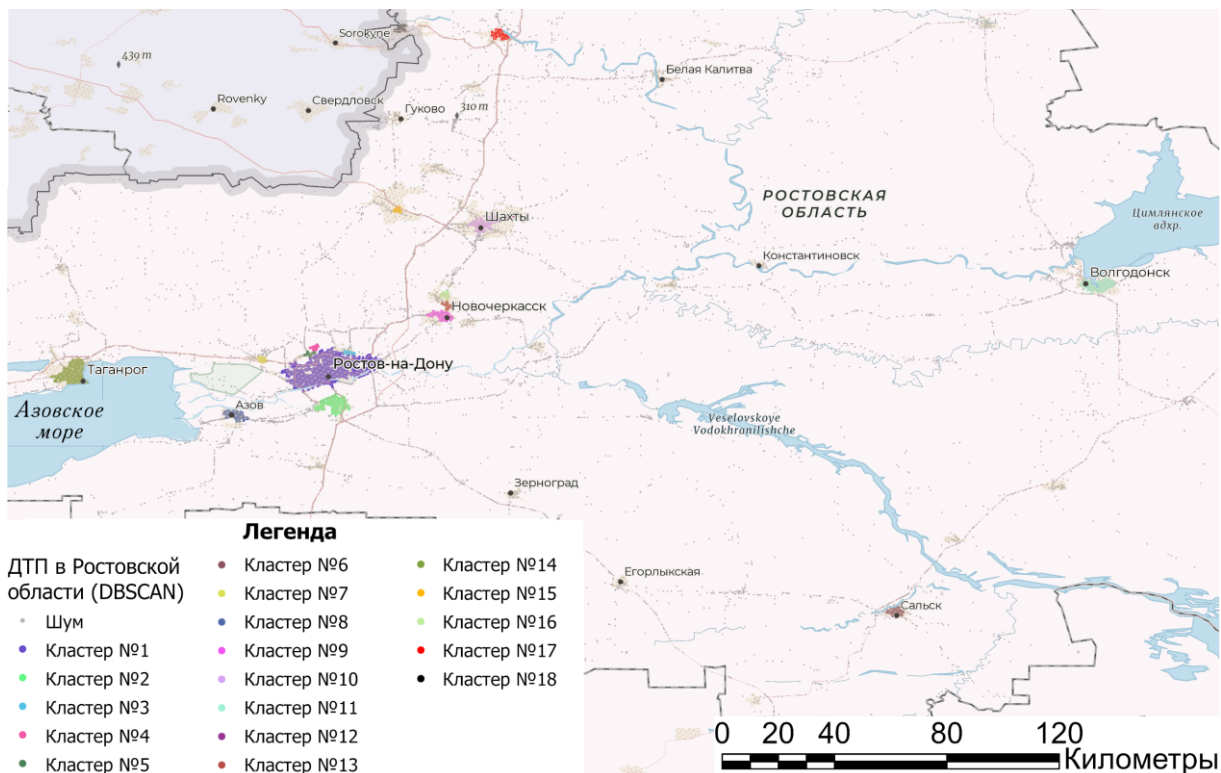


Рис. 2. Результат работы алгоритма DBSCAN. Вариант 1  
 Fig.2. The result of the DBSCAN algorithm. Option 1



Рис. 3. Результат работы алгоритма DBSCAN. Вариант 2  
 Fig. 3. The result of the DBSCAN algorithm. Option 2

Особенностью работы алгоритма HDBSCAN является автоматический подбор расстояний поиска, что, с одной стороны, делает данный алгоритм наиболее удобным в использовании, а с другой стороны, создает большое количество несопоставимых по размерам и формам кластеров. Это делает проблематичным их дальнейший анализ. Полученные кластеры представлены на рис. 4–5.

В отличие от предыдущих алгоритмов, OPTICS при построении кластеров использует указанное пользователем значение кластерной чувствительности (от 0 до 100, в работе использовалось значение 100), характеризующее плотность в построенных кластерах. При значениях варианта 1 алгоритм построил кластеры аварийности в следующих городах Ростовской области: Таганрог, Ростов-на-Дону, Батайск, Чалтырь, Азов, Новочеркасск, Шахты, Донецк, Каменск-Шахтинский, Волгодонск, Сальск. Кластеры, построенные с помощью алгоритма OPTICS при значениях варианта 1, приведены на рис. 6.

При значениях варианта 2 алгоритм построил кластеры аварийности также на самых оживленных участках. Некоторые из них совпали с алгоритмом DBSCAN, однако часть предыдущих кластеров была проигнорирована. Объясняется это тем, что OPTICS обнаруживает кластеры примерно одинаковой плотности, что крайне важно в условиях анализа больших данных. Без этого свойства есть риск того, что обнаруженные кластеры будут иметь много статистически незначимых объектов ДТП, из-за чего точность выявления будет снижена. Результат работы алгоритма OPTICS при значениях варианта 2 представлен на рис. 7.

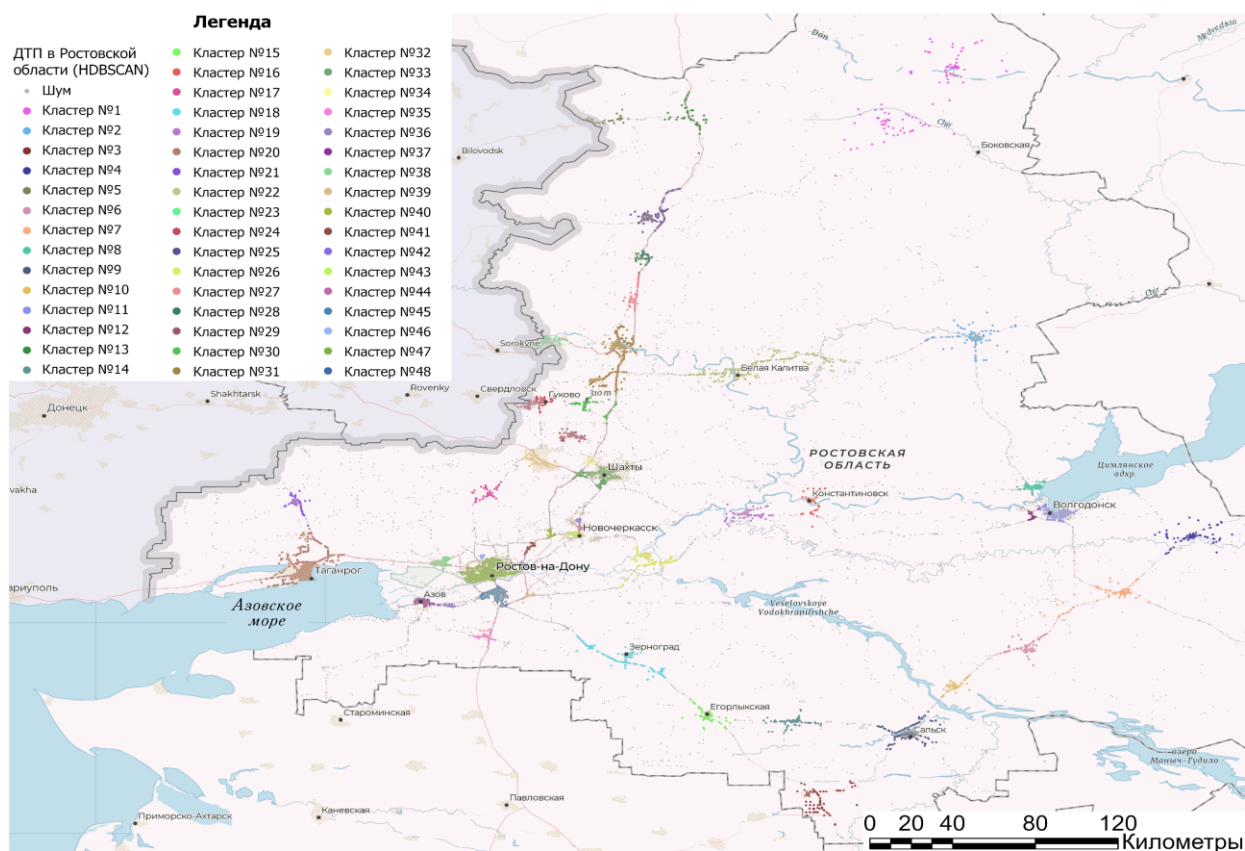


Рис. 4. Результат работы алгоритма HDBSCAN. Вариант 1

Fig. 4. The result of the HDDSCAN algorithm. Option 1

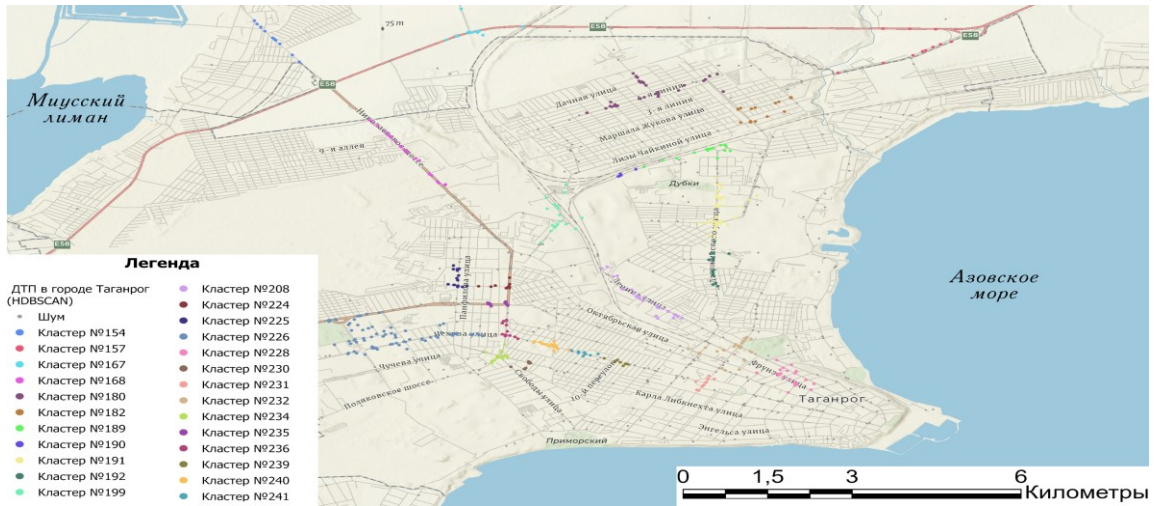


Рис. 5. Результат работы алгоритма HDBSCAN. Вариант 2  
 Fig. 5. The result of the HDDSCAN algorithm. Option 2

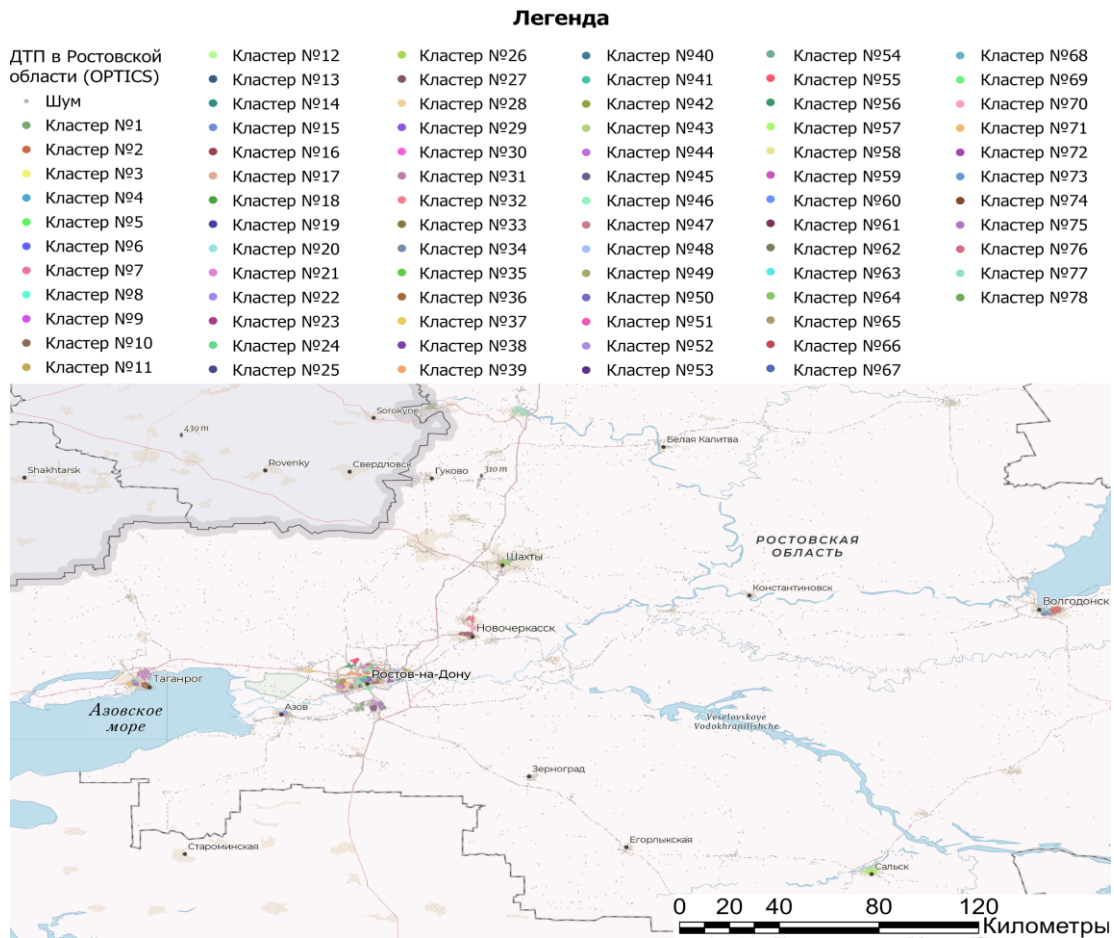


Рис. 6. Результат работы алгоритма OPTICS. Вариант 1  
 Fig. 6. The result of the OPTICS algorithm. Option 1



Рис. 7. Результат работы алгоритма OPTICS. Вариант 2  
 Fig. 7. The result of the OPTICS algorithm. Option 2

Этап 2. Анализ построенных кластеров

Для более подробного анализа были взяты кластеры, построенные алгоритмом OPTICS «Вариант 2» для г. Таганрога. Их местоположение представлено в табл. 1.

Табл. 1. Местоположение кластеров OPTICS  
 Table 1. Location of OPTICS clusters

Номер кластера	Местоположение
219	пересечении трасс А-280 и 60к-232
235	пересечение улиц Петровской и Социалистической
236	пересечение улицы Чехова и Большого проспекта
237	пересечение улицы Чехова и 16-го переулка
238	пересечение улиц Театральной и Транспортной

Самыми частыми категориями ДТП для всех кластеров является «наезд на пешехода/велосипедиста» и «столкновение». Так, в 236 кластере за период с 2015 по 2022 гг. 56 % ДТП связаны с наездом на пешехода, а в 237 кластере 90 % ДТП являются столкновением транспортных средств. Процентное соотношение категорий ДТП в исследуемых кластерах представлено на рис. 8.



Рис. 8. Диаграмма ДТП по категориям  
 Fig. 8. Accident diagram by category

В 3 из 5 кластеров в зоне ДТП была в наличии горизонтальная дорожная разметка, что отражено на рис. 9. В остальных кластерах разметка отсутствовала вследствие изношенности либо плохих погодных условий.



*Рис. 9. Диаграмма ДТП по наличию/отсутствию горизонтальной разметки*

*Fig. 9. Diagram of an accident by the presence/absence of horizontal markings*

На основании атрибутивных данных был сделан вывод о том, что основной причиной ДТП в кластерах является «непредоставление преимущества на перекрестке», поэтому для снижения аварийности рекомендуется в данных кластерах внести изменения в схему организации дорожного движения.

Дальнейший анализ данных и исследование территорий позволят более точно определить, какие характеристики сегментов участков дорожной сети приводят к высокой концентрации ДТП.

## ВЫВОДЫ

В данной работе показано, что геоинформационный метод выявления и анализа, с одной стороны, способен собрать и отобразить всю необходимую для анализа информацию, а с другой стороны, при использовании методов машинного обучения и, в частности, кластеризации, автоматизировать все процессы, связанные с обработкой данных в ГИС. По результатам проведенного сравнения алгоритмов DBSCAN, HDBSCAN, OPTICS можно сказать следующее: для задач пространственного анализа ДТП при увеличении количества числа больших данных высокой плотности DBSCAN показывает меньшую эффективность, чем OPTICS, из-за различий в плотности в обнаруживаемых кластерах. Использование HDBSCAN не исправляет данного недостатка, т. к. подбираемое в результате работы динамическое расстояние между объектами создает большое количество разнородных кластеров. В свою очередь, OPTICS позволяет обнаруживать кластеры одинаковых размеров и плотности, что делает его наиболее подходящим алгоритмом для получения стабильных и содержательных кластеров ДТП.

Дальнейшие исследования позволят определить статистическую достоверность построенных кластеров, а также эффективность кластеризации при увеличении масштабов территории анализа. Использование ГИС, данных дистанционного зондирования, спутниковой навигации для изучения скоростей движения позволит в конечном итоге добиться создания комплексного решения, способного повысить эффективность выявления и анализа УКДТП.

## СПИСОК ЛИТЕРАТУРЫ

*Герштейн А. М., Терехов А. Н.* Выявление участков повышенной опасности на дорогах Массачусетса в 2013–2018 годах. Компьютерные инструменты в образовании, 2021. № 1(19). С. 46–58. DOI: 10.32603/2071-2340-2021-1-46-58.

*Елисеев М. Е., Галкина Е. Д., Мазунова Л. Н.* Критерий для выявления мест концентрации аварий при разработке интеллектуальной информационной транспортной системы. Труды НГТУ им. Р. Е. Алексеева, 2021. № 1. С. 14–23. DOI: 10.46960/1816-210X\_2021\_1\_14.

- Al-Haji G.* Traffic safety in developing countries — new approaches in technology transfer by using distance education technique. Master's thesis, LITH-ITN-2001:156-SE, Linköping University, Sweden, 2001.
- Arjun D.* Clustering techniques and their applications: a review. American journal of advanced computing, 2020. V. 1. No. 4. P. 1–6. DOI: 10.15864/ajac.1404.
- Nisha, Kaur P. J.* A survey of clustering techniques and algorithms. 2015 2nd International conference on computing for sustainable global development (INDIACom), 2015. P. 304–307.
- Prasad V., Srikanth T.* A survey on clustering algorithms and their constraints. International journal of intelligent systems and applications in engineering, 2023. No. 11(6s). P. 165–179.
- Smeed R. J.* Some statistical aspects of road safety research. Journal of the royal statistical society. Series A (general), 1949. V. 112. No. 1. P. 1–34.
- Soloviev A. A., Zharkikh J. I., Krasnoperov R. I., Nikolov B. P., Agayan S. M.* GIS-oriented solutions for advanced clustering analysis of geoscience data using ArcGIS platform. Russian journal of Earth sciences, 2016. V. 16. DOI: 10.2205/2016ES000587.
- Haoxiang Xu.* Research on clustering algorithms in data mining. 2022 3rd International conference on big data, artificial intelligence and internet of things engineering (ICBAIE), 2022. P. 652–655. DOI: 10.1109/ICBAIE56435.2022.9985831.

#### REFERENCES

- Al-Haji G.* Traffic safety in developing countries — new approaches in technology transfer by using distance education technique. Master's thesis, LITH-ITN-2001:156-SE, Linköping University, Sweden, 2001.
- Arjun D.* Clustering techniques and their applications: a review. American journal of advanced computing, 2020. V. 1. No. 4. P. 1–6. DOI: 10.15864/ajac.1404.
- Gershtein A. M., Terekhov A. N.* Identification of high-risk areas on Massachusetts roads in 2013–2018. Computer tools in education Journal, 2021. No. 1 (19). P. 46–58 (in Russian). DOI: 10.32603/2071-2340-2021-1-46-58.
- Eliseev M. E., Galkina E. D., Mazunova L. N.* Criterion for identifying accident concentration areas when developing an intelligent information transport system. Proceedings of NNSTU n. a. R. E. Alekseev, 2021. No. 1. P. 14–23 (in Russian). DOI: 10.46960/1816-210X\_2021\_1\_14.
- Nisha, Kaur P. J.* A survey of clustering techniques and algorithms. 2015 2nd International conference on computing for sustainable global development (INDIACom), 2015. P. 304–307.
- Prasad V., Srikanth T.* A survey on clustering algorithms and their constraints. International journal of intelligent systems and applications in engineering, 2023. No. 11(6s). P. 165–179.
- Smeed R. J.* Some statistical aspects of road safety research. Journal of the royal statistical society. Series A (general), 1949. Vol. 112. No. 1. P. 1–34.
- Soloviev A. A., Zharkikh J. I., Krasnoperov R. I., Nikolov B. P., Agayan S. M.* GIS-oriented solutions for advanced clustering analysis of geoscience data using ArcGIS platform. Russian journal of Earth sciences, 2016. V. 16. DOI:10.2205/2016ES000587.
- Haoxiang Xu.* Research on clustering algorithms in data mining. 2022 3rd International conference on big data, artificial intelligence and internet of things engineering (ICBAIE), 2022. P. 652–655. DOI: 10.1109/ICBAIE56435.2022.9985831.