

УДК 004.891.2:681.518  
DOI

## РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ ПО ПОДБОРУ ОЗДОРОВИТЕЛЬНОГО ОТДЫХА В РЕГИОНЕ КAVКАЗСКИХ МИНЕРАЛЬНЫХ ВОД НА ОСНОВЕ АНАЛИЗА МЕДИЦИНСКИХ ПРОФИЛЕЙ И СТОИМОСТНЫХ ХАРАКТЕРИСТИК

<sup>1</sup>Ковалева А.О., <sup>1</sup>Мартirosян К.В., <sup>1</sup>Цаплева В.В., <sup>2</sup>Суюнова Г.Б.

<sup>1</sup>Федеральное государственное автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет», Пятигорск, Российская Федерация, e-mail: belycka@mail.ru;

<sup>2</sup>Федеральное государственное бюджетное образовательное учреждение высшего образования «Пятигорский государственный университет», Пятигорск, Российская Федерация

В рамках данной работы представлена разработка интеллектуальной рекомендательной системы. Система использует методы обработки естественного языка и машинного обучения для персонализированного подбора программы лечения в санаторно-курортном учреждении Кавказских Минеральных Вод. Методы обработки основаны на анализе медицинских профилей, уровне сервиса и ценовых показателях объекта исследования. Цель исследования – повышение эффективности работы алгоритмов выбора санатория путем разработки автоматизированной рекомендательной системы по подбору оздоровительного отдыха. Материалы исследования и методы включают обработку структурированных данных, ранжирование объектов по заданным критериям и визуализацию результатов с использованием методов кластеризации. Представленные методы снижают временные затраты на поиск подходящего санатория за счет обработки большого массива информации с применением интеллектуальных технологий. Результаты и их обсуждение позволяют определить конфигурацию интерфейса приложения, а также подтвердить предположение о возможности кластеризации применительно к параметрам объекта. Результаты работы определяют возможность применения интеллектуальных технологий кластеризации для определения признаков объекта исследования с целью разработки эффективного приложения. Методы визуализации, примененные в ходе разработки программного продукта, позволяют пользователю получить информацию об объекте интуитивно понятным образом.

**Ключевые слова:** рекомендательная система, машинное обучение, медицинский туризм, кластеризация данных, Python, обработка структурированных данных, оценка сходства через косинусную меру

**Благодарности.** Автор благодарит профессора Першина И.М. за содействие в исследовании системы управления.

## DEVELOPMENT OF AN AUTOMATED RECOMMENDATION SYSTEM FOR SELECTION OF HEALTH RECREATION IN THE REGION OF THE CAUCASIAN MINERAL WATERS BASED ON ANALYSIS OF MEDICAL PROFILES AND COST CHARACTERISTICS

<sup>1</sup>Kovaleva A.O., <sup>1</sup>Martirosyan K.V., <sup>1</sup>Tcapleva V.V., <sup>2</sup>Suyunova G.B.

<sup>1</sup>Federal State Autonomous Educational Institution of Higher Education “North-Caucasus Federal University (NCFU)”, Pyatigorsk, Russian Federation, e-mail: belycka@mail.ru;

<sup>2</sup>Federal State Budgetary Educational Institution of Higher Education “Pyatigorsk State University (PGU)”, Pyatigorsk, Russian Federation

This paper presents the development of an intelligent recommendation system. The system utilizes natural language processing and machine learning to provide personalized treatment program selection at a health resort in the Caucasian Mineral Waters region. The methods are based on an analysis of medical profiles, service levels, and price indicators for the subject. The objective of the study was to improve the efficiency of resort selection algorithms by developing an automated recommendation system for health vacation selection. The research materials and methods include processing structured data, ranking objects according to specified criteria, and visualizing the results using clustering methods. The presented methods reduce the time spent searching for a suitable resort by processing a large array of information using intelligent technologies. The results and their discussion allow us to determine the configuration of the application interface and confirm the hypothesis regarding the feasibility of clustering with respect to the object's parameters. The results of the study demonstrate the feasibility of using intelligent clustering technologies to identify the features of the subject for the purpose of developing an effective application. The visualization methods used during the development of the software product allow the user to obtain information about the object in an intuitive manner.

**Keywords** recommender system, machine learning, medical tourism, data clustering, Python, structured data processing, similarity assessment through the cosine measure

**Acknowledgements.** The author thanks Professor I.M. Pershin for his assistance in researching the control system.

### Введение

Роль систем управления в технических науках невозможно переоценить. Применение технологий машинного обучения позволяет использовать структуры систем управления на новом уровне. В данной работе показано, как с помощью инструментов Python и библиотеки ScikitLearn можно эффективно выполнять разработку автоматизированной рекомендательной системы. Особенно важной является возможность настройки сервисов системы непосредственно в процессе разработки.

Предметной областью исследования выбраны курортно-оздоровительные технологии. Большой спектр выбора методов санаторно-курортного лечения наводит на мысль о необходимости разработки рекомендательной системы по подбору оздоровительного отдыха. Одним из важнейших компонентов для реабилитации и поддержания здоровья каждого человека является санаторно-курортное лечение, основанное на использовании природных лечебных факторов. Санаторный отдых в условиях пребывания в лечебно-оздоровительной местности эффективно и целенаправленно воздействует на врожденные или приобретенные заболевания, в том числе и на ряд хронических заболеваний. Рост спроса на медицинский туризм в России требует создания автоматизированных систем, способных анализировать параметры санаториев [1].

При выборе санатория необходимо учитывать такие факторы, как медицинский профиль заведения, географическое положение, уровень сервиса и стоимости проживания на одного человека [2]. Существующие платформы бронирования часто ограничиваются фильтрами по базовым параметрам, не предлагая персонализированных рекомендаций [3].

**Цель исследования** – повышение эффективности работы алгоритмов выбора санатория путем разработки автоматизированной рекомендательной системы по подбору оздоровительного отдыха.

Задачей данной работы ставится разработка интеллектуального алгоритма. Алгоритм на основе анализа медицинских показаний и стоимостных характеристик автоматически подбирает подходящий санаторий и визуализирует результат для сравнительного анализа. Внедрение такой системы позволит пациентам сократить время поиска и повысить качество выбора. Санатории увеличат заполняемость и снизят количество отказов.

### Материалы и методы исследования

Актуальность разработки определяется внедрением экспертно-рекомендательной системы. Такая система сочетает анализ данных, искусственный интеллект и предметные знания [4; 5]. Применение сервисов машинного обучения не только упрощает выбор санатория под индивидуальные запросы пользователя, но и задает новые стандарты в развитии технологий в области курортной медицины [6].

Для реализации системы выбрана среда GoogleColab, язык программирования Python, библиотеки Pandas, Scikit-learn и Matplotlib.

В рамках исследования использован специализированный датасет, отрывок которого представлен на рисунке 1. Датасет содержит 25 наименований санаториев Кавказских Минеральных Вод. В датасете включены следующие параметры санаториев: медицинские профили (номинальные переменные), уровень сервиса (2–5 баллов) и минимальная цена проживания (руб./сутки).

Санаторий	Город	Профиль 1	Профиль 2	Профиль 3	Профиль 4	Профиль 5	Уровень серви...	Цена за сутки...
Плаза	Железноводск	Желудочно-к...	Гинекология	Обмен веществ	Урология	Печень и жел...	4	от 8500
им 30 лет Поб...	Железноводск	Гинекология	Мочеполовая...	Урология	Сердечно-сос...	Желудочно-к...	4	от 5450
Машук Аква-Т...	Железноводск	Гинекология	Дыхательная...	Кожные забол...	ЛОР	Обмен веществ	4	от 9800
Элита	Кисловодск	Дыхательная...	Имунная сист...	Кровообраще...	ЛОР	Нервная сист...	4	от 4200
Москва	Кисловодск	Дыхательная...	Кровообраще...	Нервная сист...	Сердечно-сос...	Обмен веществ	3	от 3900
Пятигорский...	Пятигорск	Гинекология	Кожные забол...	Нервная сист...	Опорно-двиг...	Урология	4	от 6748
Виктория	Ессентуки	Имунная сист...	Обмен веществ	Эндокринная...	Дыхательная...	Кровообраще...	3	от 3000
Солнечный	Кисловодск	Дыхательная...	Кровообраще...	ЛОР	Нервная сист...	Сердечно-сос...	4	от 7650
Зори Ставроп...	Пятигорск	Кожные забол...	Мочеполовая...	Урология	Опорно-двиг...	Гинекология	3	от 5700
Долина Нарза...	Кисловодск	Дыхательная...	Кровообраще...	Нервная сист...	Опорно-двиг...	Сердечно-сос...	4	от 8600
Дон	Пятигорск	Желудочно-к...	Мочеполовая...	Печень и жел...	Урология	Дыхательная...	3	от 4000
Дубрава	Железноводск	Мочеполовая...	Педиатрия	Урология	Эндокринная...	Печень и жел...	3	от 3800
Русь	Ессентуки	Нервная сист...	ЛОР	Сердечно-сос...	Эндокринная...	Дыхательная...	4	от 9600
Целебный ключ	Ессентуки	Обмен веществ	Печень и жел...	Похудение/О...	Эндокринная...	Имунная сист...	2	от 3200
Славяновский...	Железноводск	Желудочно-к...	Мочеполовая...	Урология	Нервная сист...	Имунная сист...	4	от 6950
Машук	Пятигорск	Гинекология	Офтальмология	Эндокринная...	Мочеполовая...	Опорно-двиг...	3	от 5240

Рис. 1. Фрагмент представления данных, используемых при решении задачи  
Источник: составлено авторами на основе исследования

Ключевые функциональные элементы включают в себя интерактивный интерфейс для выбора заболеваний, кластеризацию объектов по двум параметрам (цена и рейтинг обслуживания) и автоматизированный подбор аналогов на основе семантического анализа профилей лечения [7].

Система рекомендаций состояла из двух этапов: фильтрации и ранжирования [8; 9]. Пользователь взаимодействует с системой через выпадающий список. Опции формируются автоматически из столбца с профилями заболеваний загруженного датасета.

Кодирование текстовых данных, по которым система выполняет фильтрацию, реализуется посредством перевода текстового списка в вектор-столбец данных  $P_i$ , как показано в формуле (1.1). Аналогичным образом обрабатывается список заболеваний.

Для заданного пользователем списка заболеваний  $D=\{d_1, d_2, \dots, d_n\}$  система формирует булеву маску, в которой осуществляется поиск санаториев, где хотя бы один профиль  $p_j$  совпадает с  $d_i$  заболеванием:

$$\text{mask} = \bigcup_{i=1}^n \left( \bigcup_{j=1}^5 (p_j \equiv d_i) \right)$$

где  $p_j$  – медицинский профиль санатория.

Данное условие позволяет системе выбрать санатории, которые специализируются хотя бы на одном из отмеченных пользователем заболеваний. Это позволяет предоставлять рекомендации даже при неполном совпадении профилей, что критически важно для пациентов с редкими диагнозами либо с комплексными патологиями (например, сочетание дерматологических и кардиологических проблем). При этом

для редких заболеваний система предлагает варианты лечения даже в том случае, если ни один санаторий полностью не соответствует запросу. В случае сочетанных заболеваний система найдет учреждения, где лечат хотя бы одну из этих проблем.

Для решения данной задачи использовался алгоритм К-средних (K-Means) из библиотеки ScikitLearn [9]. Основной особенностью реализации была кластеризация по признакам (уровень сервиса и стоимость проживания), а также определение количества кластеров от 1 до 4 в зависимости от размера выборки.

Именно метод К-средних идеально отражает зависимость между гибкостью критериев, динамическим обновлением и удобством ввода для обычного обывателя [10; 11]. На рисунке 2 изображен принцип использования данного метода.

Для предложения альтернатив использовался метод косинусного сходства, оценивающий семантическую близость текстовых профилей. Данный метод является достаточно эффективным для сравнения текстовых профилей небольшой длины [12]. Основное преимущество применения косинусного сходства – инвариантность к масштабу. Сходство между векторами разной длины, но одного направления будет высоким [13; 14].

Критерий выбора количества кластеров определяется технологией метода k-средних. При этом количество кластеров задается при запуске алгоритма обработки и может меняться при необходимости. В данном случае оптимальным количеством с точки зрения визуализации результата является разбиение на четыре кластера.

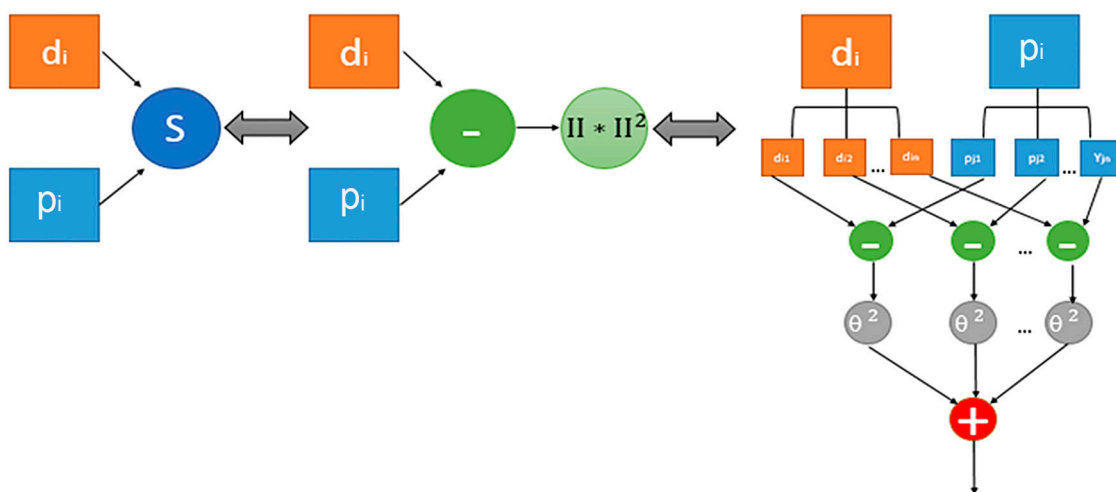


Рис. 2. Метод k-средних

Источник: составлено авторами на основе данного исследования

Оценка качества алгоритма кластеризации была проведена с применением метрики качества «коэффициент Силуэтта». Такая оценка позволяет понять, насколько релевантна полученная выборка. Метрика оценивает качество алгоритма на основе сравнения условного расстояния между объектами кластера и между объектами разных кластеров. Качество кластеризации может быть оценено как высокое, если результат расчета метрики близок к единице. Расчеты коэффициента Силуэтта показали релевантность применяемого метода. Коэффициент Силуэтта для данного исследования составляет 0,87.

Метод кластеризации представляет собой оптимальный баланс между точностью, скоростью и интерпретируемостью [15]. Обработка данных с применением этого метода показала высокую эффективность при решении поставленной задачи

и позволила сформировать интерфейс программного продукта.

### Результаты исследования и их обсуждение

Важным этапом анализа результатов кластеризации является визуализация вывода. Метод визуализации позволяет показать пользователю выборку санаториев на основе применяемой методики. На рисунке 3 представлен график результата кластеризации санаториев по цене и уровню сервиса, где размер и цвет маркеров делают выводы более наглядными, а надпись соответствует названию санатория.

При построении графика применяется принцип многомерного представления информации, где конкретно каждый тип данных кодируется отдельным свойством с визуальными критериями [16; 17]. Это позволяет без потери читаемости совместить на одном графике несколько измерений.

Карта санаториев по цене и уровню сервиса



Рис. 3. График кластеризации по цене и уровню сервиса  
Источник: составлено авторами на основе данного исследования



Размер маркеров показывает стоимость. Логично, что чем больше санаторий, тем он дороже. Это интуитивно понятно, так как люди обычно соотносят размер с величиной. Тем самым пользователь может сразу отфильтровать маркеры на основе своего бюджета.

Цвет маркеров соответствует городу. Это помогает группировать санатории по географическому признаку. Если пользователь предпочитает определенный регион, он сразу может обратить внимание на маркеры нужного цвета. Важно подчеркнуть, что использование разных цветов упрощает идентификацию и сравнение между кластерами, тем самым пользователь может выбрать в од-

ном городе схожие по сервису, но разные по цене санатории.

Подписи с названиями санаториев необходимы для точной идентификации каждого пункта на графике. Они работают как «якоря», связывая визуальные элементы с текстовой информацией [18]. Даже при близком расположении точек пользователь точно идентифицирует каждый санаторий. Для улучшения восприятия графика подобран шрифт, а также выполнено выравнивание по правому краю, предотвращающее наложение текста. Данный подход позволяет устранить хаотичность картины, минимизируя накладки надписей друг на друга. Это позволяет избежать пересечения информационных объектов [19].

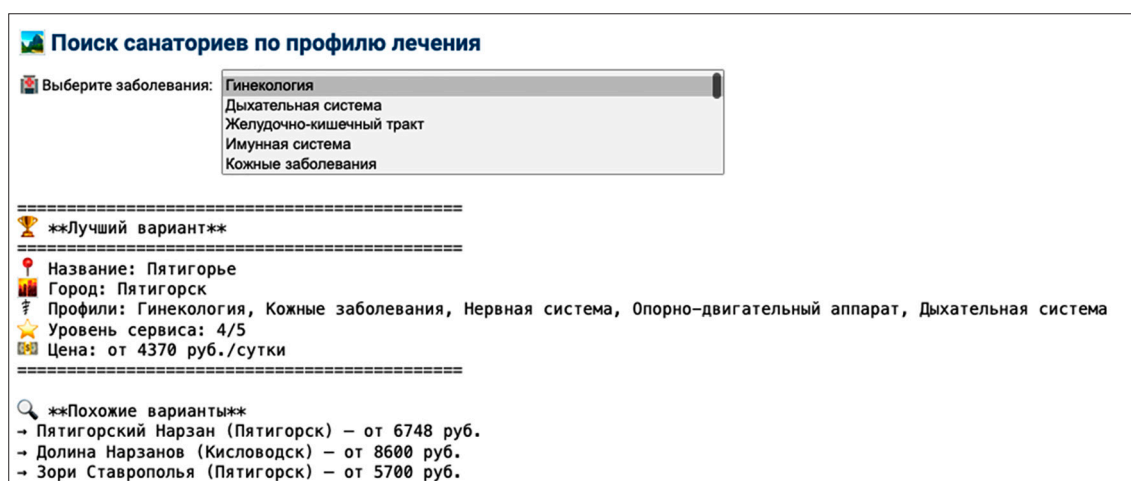


Рис. 4. Интерактивный интерфейс при выборе пользователем профиля лечения «Гинекология»  
Источник: составлено авторами на основе данного исследования

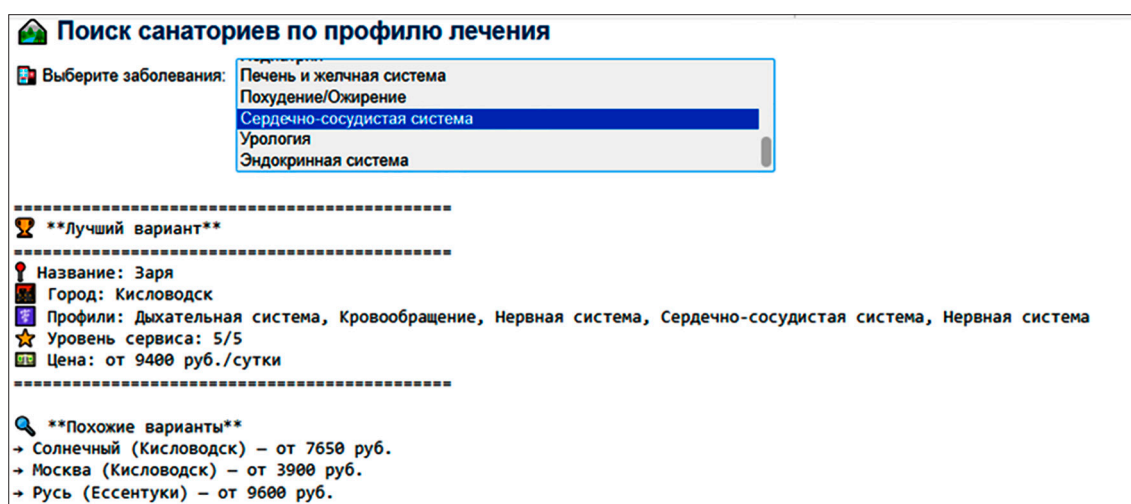


Рис. 5. Интерактивный интерфейс при выборе пользователем профиля лечения «Сердечно-сосудистая система»  
Источник: составлено авторами на основе данного исследования

Спектр методов визуализации обширен и не ограничивается предлагаемой картиной. Такой популярный инструмент, как тепловые карты, в данном случае неэффективен, так как система работает с дискретными данными (уровень сервиса – целые числа). Применение столбчатых диаграмм затруднит визуальный анализ большого количества объектов по нескольким категориям.

Предлагаемая методика визуализации наиболее эффективна в данной постановке задачи. Сочетание цвета, размера и текстовых меток маркеров на точечном графике формирует информационно-эффективный, а также интуитивно понятный интерфейс для обычного пользователя. Данный выбор сделан на основе анализа альтернативных способов визуализации (тепловые карты, столбчатые и круговые диаграммы) и соответствует задаче сравнения объектов по двум ключевым параметрам [20].

Интерактивный интерфейс представлен скриншотами программы. Рассматривается случай, когда пользователь выбирает профиль лечения «Гинекология» (рис. 4). Еще один скриншот показывает процесс анализа при выборе профиля «Сердечно-сосудистая система» (рис. 5).

Для взаимодействия с пользователем используется виджет SelectMultiple из библиотеки `ipywidgets`. Динамическое изменение критериев позволяет пользователю экспериментировать с разными комбинациями заболеваний и сразу видеть, как меняются рекомендации [21]. Эта интерактивность делает систему более гибкой и отзывчивой, что улучшает пользовательский опыт.

### Заключение

Данная разработка не только упрощает навигацию в выборе лечебного учреждения, но и повышает точность рекомендаций за счет исключения человеческого фактора. Предлагаемая методика минимизирует риски некорректного подбора терапии, что критично для эффективной реабилитации.

В данной работе демонстрируются возможности применения искусственного интеллекта в разработке рекомендательных систем. Тестирование алгоритмов анализа данных и работа с интерфейсом приложения показали релевантность предлагаемого метода. Автоматизированная рекомендательная система по подбору оздоровительного отдыха является экспертной системой с базой знаний. Такую систему можно доработать в следующей итерации до формата интеллектуального агента-советчика. Формат генератора рекомендаций по профилю лечения является своевременным предложением.

Основной проблемой разработки такого генератора является состав определяющих решение пользователя параметров санаторно-курортного отдыха. Решение проблемы заключается в применении технологий кластеризации. Искусственный интеллект определяет эффективный набор признаков, которые будут использованы в качестве фильтров выбора санатория. Расчет метрик качества кластеризации на основе коэффициента Силуэтта показал релевантность применяемых методов.

Технологии искусственного интеллекта позволили сформировать эргономичный и интуитивно доступный интерфейс программного приложения. Пользователь получает доступ к выбору профиля лечения в санатории, ему предоставляется визуальная картина результатов выбора. Таким образом, технологии машинного обучения позволяют существенным образом упростить процессы анализа данных и непосредственно разработку программных приложений.

### Список литературы

1. Ботвинева Н.Ю., Браткова В.В., Давтян А.А. Особенности стратегического управления в формировании стратегического имиджа города-курорта Пятигорска // Вестник экспертного совета. 2025. № 1 (40). [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/osobennosti-strategicheskogo-upravleniya-v-formirovani-imidzha-goroda-kurorta-pyatigorska> (дата обращения: 02.10.2025).
2. Жукова А.Е., Куров А.Е. К вопросу об актуализации стратегии социально-экономического развития города Пятигорска // Вестник экспертного совета. 2025. № 1 (40). [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/k-voprosu-ob-aktualizatsii-strategii-sotsialno-ekonomicheskogo-razvitiya-goroda-pyatigorska> (дата обращения: 02.10.2025).
3. Лихтин Д.А. Санаторно-курортные услуги: природа, специфика и подходы к классификации // Экономика и управление. 2022. № 9. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/sanatorno-kurortnye-uslugi-priroda-spetsifika-i-podhody-k-klassifikatsii> (дата обращения: 02.10.2025).
4. Першин И.М., Помеляйко И.С. Системный анализ экологического состояния зоны гипергенеза курорта Кисловодска // Вестник Северо-Кавказского федерального университета. 2013. № 3 (36). С. 74-80. [Электронный ресурс]. URL: <https://www.elibrary.ru/item.asp?id=21031649> (дата обращения: 02.10.2025).
5. Першин И.М., Цаплева В.В., Носова В.А. Верификация дискретных моделей гидротермальных процессов // Современная наука и инновации. 2021. № 4 (36). С. 31-39. URL: <https://msi.elpub.ru/jour/article/view/470/470?ysclid=mg96r61d7399592678> (дата обращения: 02.10.2025).
6. Першин И.М., Носова В.А., Русак С.Н. Определение параметров распределенных звеньев, аппроксимирующих гидротермальные процессы // Современная наука и инновации. Научный журнал. 2021. Вып. № 2 (34). С. 10-16. URL: <https://elibrary.ru/item.asp?id=47415268&ysclid=mg96tlfp93506808598> (дата обращения: 02.10.2025).
7. Першин И.М., Веселов Г.Е. Аппроксимационные модели передаточных функций распределенных объектов // Известия ЮФУ. Технические науки. 2015. № 7 (168). [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/approksimatsionnye-modeli-peredatochnykh-funktsiy> (дата обращения: 02.10.2025).

8. Першин И.М., Цаплева В.В., Носова В.А. Математическое моделирование гидролитосферных процессов с учетом процессов в «колодцах» // Современная наука и инновации. 2022. № 3 (39). С. 54-61. DOI: 10.37493/2307-910X.2022.3.5.
9. Першин И.М., Цаплева В.В., Носова В.А. Самонастраивающиеся распределенные системы управления // Современная наука и инновации. 2023. № 1 (41). С. 15-28. DOI: 10.37493/2307-910X.2023.1.2.
10. Pershin I.M., Papush E.G., Kukharova T.V., Utkin V.A. Modeling of Distributed Control System for Network of Mineral Water Wells // Water. 2023. Vol. 15. № 12. DOI: 10.3390/w15122289.
11. Першин И.М., Малков А.В., Веселов Г.Е., Першин М.И. Автоматизация технологических процессов добычи гидроминерального сырья // Известия ЮФУ. Технические науки. 2018. № 5 (199). [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/avtomatizatsiya-tehnologicheskikh-protsessov-dobychi-gidromineralnogo-syrya> (дата обращения: 02.10.2025).
12. Ilyushin Y.V., Kapostey E.I. Developing a Comprehensive Mathematical Model for Aluminium Production in a Soderberg Electrolyser // Energies. 2023. № 16. 6313. DOI: 10.3390/en16176313.
13. Asadulagi M.A., Pershin I.M., Tsapleva V.V. Research on Hydrolithospheric Processes Using the Results of Groundwater Inflow Testing // Water. 2024. Vol. 16. № 3. DOI: 10.3390/w16030487.
14. Afanaseva O., Bezyukov O., Pervukhin D., Tukeev D. Experimental Study Results Processing Method for the Marine Diesel Engines Vibration Activity Caused by the Cylinder-Piston Group Operations // Inventions. 2023. № 8 (3). P. 71. DOI: 10.3390/inventions8030071.
15. Martirosyan A.V., Ilyushin Y.V. The Development of the Toxic and Flammable Gases Concentration Monitoring System for Coalmines // Energies. 2022. Vol. 15. № 23. P. 8917. DOI: 10.3390/en15238917.
16. Першин И.М. Исследование характеристик пространственных фильтров // Современная наука и инновации. 2024. № 2 (46). С. 48-58. DOI: 10.37493/2307-910X.2024.2.5.
17. Георгиева М.А., Першин И.М. Математическое моделирование взаимовлияния атмосферных осадков и гидролитосферных процессов на наклонные поверхности // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 1. С. 133-142. DOI: 10.35330/1991-6639-2025-27-1-133-142.
18. Першин И.М., Носова В.А. Исследование распределенных объектов и систем с использованием пространства  $\{G, RE, IM\}$  // Современная наука и инновации. 2024. № 3 (47). С. 50-64. DOI: 10.37493/2307-910X.2024.3.5.
19. Martirosyan A.V., Ilyushin Y.V., Afanaseva O.V. Development of a Distributed Mathematical Model and Control System for Reducing Pollution Risk in Mineral Water Aquifer Systems // Water. 2022. Vol. 14. № 2. DOI: 10.3390/w14020151.
20. Martirosyan A.V., Ilyushin Y.V. Modeling of the Natural Objects' Temperature Field Distribution Using a Supercomputer // Informatics. 2022. Vol. 9. № 3. P. 62. DOI: 10.3390/informatics9030062.
21. Георгиева М.А., Першин И.М. Математическое моделирование влияния атмосферных осадков на гидролитосферные процессы // Известия ЮФУ. Технические науки. 2024. № 6 (242). С. 121-131. DOI: 10.18522/2311-3103-2024-6-121-131.

**Конфликт интересов:** Автор заявляет об отсутствии конфликта интересов.

**Conflict of interest:** The author declares that he has no conflict of interest.