

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ / ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

DOI: <https://doi.org/10.60797/COMP.2025.5.6>

ПРОБЛЕМАТИКА АНАЛИЗА И ОЦЕНКИ РИСКОВ ЗДОРОВЬЮ НАСЕЛЕНИЯ НА БАЗЕ ИСПОЛЬЗОВАНИЯ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Научная статья

Липатова А.В.^{1,*}, Потапченко Т.Д.²

¹ORCID : 0009-0006-9729-0068;

¹МТС Диджитал, Москва, Российская Федерация

²Московский технический университет связи и информатики, Москва, Российская Федерация

* Корреспондирующий автор (myagkova_anastasia[at]mail.ru)

Аннотация

В работе рассмотрены актуальные вопросы и проблемы, обуславливающие целесообразность применения методов и подходов машинного обучения для прикладных задач, связанных с анализом экологических данных. Проведен анализ современных исследований в рассмотренной области, отмечены ключевые аспекты и проблематика проведения обработки данных. Определены и описаны особенности процедур оценки рисков здоровью населения, факторы пагубного воздействия и проблемные аспекты анализа данных. Описаны преимущества, возможности и типы методов машинного обучения, в том числе с учителем, с его частичным задействованием, без учителя и с подкреплением, рассмотрена специфика использования методов для прикладных задач анализа данных. Обозначены перспективные пути совершенствования направлений и методов машинного обучения для повышения эффективности анализа экологических данных.

Ключевые слова: машинное обучение, ансамбли, искусственный интеллект.

CHALLENGES OF ANALYSIS AND EVALUATION OF PUBLIC HEALTH RISKS BASED ON THE USE OF MACHINE LEARNING METHODS

Research article

Lipatova A.V.^{1,*}, Potapchenko T.D.²

¹ORCID : 0009-0006-9729-0068;

¹MTS Digital, Moscow, Russian Federation

²Moscow Technical University of Communications and Informatics, Moscow, Russian Federation

* Corresponding author (myagkova_anastasia[at]mail.ru)

Abstract

The work examines topical questions and problems that determine the expediency of using machine learning methods and approaches for applied tasks related to the analysis of environmental data. The analysis of modern researches in the studied area is carried out, the key aspects and problems of data processing are marked. The features of the procedures of public health risk evaluation, detrimental impact factors and problematic aspects of data analysis are defined and described. Advantages, possibilities and types of machine learning methods are described, including one with a teacher, with partial involvement, without a teacher and with support, the specifics of using the methods for applied tasks of data analysis are reviewed. Promising ways of improving the directions and methods of machine learning to increase the efficiency of environmental data analysis are outlined.

Keywords: machine learning, ensembles, artificial intelligence.

Введение

В настоящее время воздействие разнообразных антропогенных факторов различной природы оказывает все большее пагубное воздействие на экологическое равновесие и общий техногенный уровень, что отрицательно сказывается, в том числе, и на здоровье человека [1]. Загрязнение химическими и биологическими отходами окружающей экосистемы из-за неэффективной и несбалансированной организации производственных процессов в различных областях человеческой деятельности, в том числе в промышленной сфере, повышает риски приобретения населением различных хронических, аллергических, бактериальных и прочих видов заболеваний, снижая прямым и косвенным образом как общий уровень иммунитета, так и среднюю продолжительность жизни [2].

В контексте данной проблемы актуальность приобретает регулярный и четко регламентированный подход оперативного мониторинга и оценки степени загрязненности, позволяющий учесть различные факторы, а также сформировать оценку риска критичности экосистемы с точки зрения нанесения вреда населению [3]. Сложность данного процесса, с одной стороны, заключается в необходимости выполнения различных измерений посредством задействования датчиков, сенсоров и аппаратных средств детекции уровня загрязненности [4]. С другой стороны, не менее важным являются процессы сбора, обработки и анализа данных, для которых характерной особенностью является большой объем и разнородность получаемой информации в разные промежутки времени. Это порождает высокую сложность выполнения данных процессов в ручном виде, в связи с чем требуется адекватная автоматизация [5], [6].

Для решения последней обозначенной задачи целесообразным является применение современных технологий, базирующихся на элементах искусственного интеллекта (ИИ), в частности, актуальным является разработка и адаптация моделей машинного обучения (МО) для интеллектуального анализа данных по различным экологически значимым факторам. Данная обоснованность обусловлена возможностью МО автоматизировать процессы поиска скрытых взаимосвязей между различными факторами в данных, в том числе посредством имплементации механизмов обучения и гибкости для сведения задач к моделям классификации, регрессии или других типовых сценариев [7].

Цель работы заключается в исследовании возможностей и особенностей использования различных методов МО для задач интеллектуального анализа экологических данных по выявлению рисков пагубного воздействия на здоровье населения.

Анализ проблематики

В контексте рассматриваемой проблематики высокую значимость как в региональных, так и в планетарных масштабах, приобретает проблема выявления и классификации состояния окружающей среды по допустимым и недопустимым уровням с точки зрения безопасной жизнедеятельности человека, а также поиск и выработка целевых мер по обеспечению процессов снижения рисков здоровья населения.

Анализируя актуальность обозначенной проблематики, следует отметить, что к ключевым категориям факторов пагубного воздействия на здоровье населения техногенного характера можно отнести:

- загрязнение атмосферного воздуха по причине поступления различных химических или биологических выбросов;
- снижение качества водных объектов, в том числе отравление обитателей рек, морей и океанов, которые затем используются населением в пищу;
- загрязнение земельных участков химическими удобрениями и промышленными отходами;
- радиационные загрязнения;
- прочие факторы [4], [7].

Количественно выполнить оценку уровня влияния негативных экологических факторов на здоровье населения достаточно сложно в ручном режиме, в том числе в связи с неоднородным характером, большим объемом данных и наличием различных трудноформализуемых признаков различного рода (социальные, производственные, антропогенные, природные аномалии и др.). Поэтому рациональным подходом для автоматизации оценки значимости оказываемого вреда загрязняющих веществ на течение и проявление человеческих заболеваний, а также для формирования прогнозных оценок рисков воздействий вредных веществ являются методы имитационного моделирования, МО и прикладные информатические технологии [5], [6].

Анализируя опыт различных зарубежных и отечественных авторов по применению МО, следует отметить ряд важных аспектов и особенностей имплементации предиктивных моделей. В частности, в статье [7] предложена модель StatEcoNet, которая сочетает графические генеративные модели статистической экологии с нейронными сетями. В работе описывается разработка гибридной модели, которая сочетает возможности статистической экологии и ИНС для моделирования распределения видов. Авторы анализируют традиционные методы моделирования распределения видов, такие как регрессионные модели, часто игнорируют ошибки в наблюдениях, что приводит к низкой точности. Путем интеграции вероятностного подхода для анализа данных о местонахождении видов и факторов среды с использованием методов МО и ИНС реализуется возможность учета сложных нелинейных взаимосвязей между факторами экологической среды и наличием определенных видов. Предложенная авторами модель протестирована на данных о распространении растений в тропических лесах, показав значительное улучшение качества предсказаний по сравнению с традиционными подходами. Т.е. предложенный в статье подход позволяет учитывать ошибки наблюдений и повышать точность прогнозирования распространения видов.

Несколько другим образом, в работе [8] авторами рассматриваются возможности глубокого обучения для анализа данных, полученных с сенсоров, камер и акустических устройств в морской экологии. Приведены результаты анализа возможностей применения глубоких ИНС для классификации и отслеживания морских организмов, а также для выявления скрытых закономерностей в шумных данных. В статье рассматривают задачи автоматизации, классификации морских организмов на подводных видео и изображениях, анализе акустических данных для изучения поведения рыб, млекопитающих, а также выявлении скрытых закономерностей в данных, собранных сенсорами (из данных о температуре, солености, уровне кислорода). Авторами установлено, что применение МО и ИНС для анализа экологических данных с подводных камер позволяет обеспечить точность классификации около 90%, что превышает показатели других подходов.

Наряду с рассмотренными работами, в исследовании [9] проводится сравнение различных алгоритмов МО, включая деревья решений, случайный лес, ИНС и метод опорных векторов, для классификации экологических данных, в частности для определения наиболее эффективных подходов для оценки состояния почвы, воздуха и качества воды. Авторы выявили, что методы Random Forest и градиентный бустинг продемонстрировали лучшие результаты в задачах классификации благодаря их способности обрабатывать большие объемы данных и учитывать нелинейные взаимосвязи, при этом метод опорных векторов оказался наиболее эффективным для небольших и сбалансированных наборов данных.

Не менее интересным и научно-практически значимым является исследование отечественного автора [10], в котором представлены результаты разработки метода прогнозирования качества воздуха в различных городах на базе МО. Исследование основано на анализе агрегированного набора данных, связанных с оценкой наличия концентрации загрязняющих веществ в рамках разных географических локаций. Согласно результатам проведенных автором экспериментов, наиболее эффективной оказалась модель случайного леса, которая учитывала временные и климатические факторы, такие как год, месяц, день, температура и влажность. Результаты данного исследования свидетельствуют о практической целесообразности применения МО для разноплановых задач анализа показателей

качества воздуха, что позволяет утверждать об эффективности и перспективности построения прогнозов изменения уровня загрязненности в будущем на базе разработанных и использованных моделей.

Таким образом, следует отметить, что результаты этапа оценки значений риска на базе применения МО могут быть использованы для идентификации приоритетности, целесообразности и срочности внедрения санитарно-гигиенических и профилактических мероприятий, целью которых является снижение негативного влияния экологической обстановки на здоровье населения.

Следует отметить, что в данном контексте оценка риска подразумевает проведение ряда вычислительных и исследовательских процедур, связанных с задачей анализа токсикологических, гигиенических и других данных (характеризующих факторы воздействия) для расчета численного значения вероятности пагубного воздействия окружающей среды на организм отдельного человека и населения в целом.

Концепции оценки рисков, используемые на практике, часто базируются на понятиях анализа расчетных значений и выработки управляющих процедур. Базовым аспектом является оценка зависимости дозы вредного фактора к ответным реакциям организма человека. Это позволяет численно оценить значение отрицательного фактора на здоровье населения, способствуя выявлению неочевидных причин появления заболеваний у людей [1], [3], [5]. Процедура проведения оценки риска здоровью населения может включать такие этапы:

- идентификация типа и масштаба опасности;
- численная оценка дозы негативного воздействия (поиск корреляции между объемом вещества и значимостью ущерба);
- оценка склонности населения к возникновению негативных факторов в здоровье;
- характеристика потенциального риска (оценка возможного объема убытков и формализация неопределенностей в процессе вычислений).

Таким образом, рациональным подходом в оценке рисков здоровью населения является МО, позволяющее обеспечить обработку и анализ данных больших объемов и разной степени структурированности, что характерно для экологического контекста [11].

Специфика МО

МО в своем современном состоянии является интегрированной научной сферой, совмещающей ряд математических, вероятностных, статистических положений, а также отдельные аспекты из теории оптимизации, дискретного моделирования и анализа, численных методов и методов обработки данных, областью применения которой часто является поиск скрытых закономерностей в информации, генерации знаний из входных данных.

В современном МО инкапсулируются разнообразные методы обучения создаваемых математико-логических моделей, в частности:

1. Аналитическое обучение, выражаемое в формировании модели МО на основе формализованных знаний экспертов, что актуально для применения в области разработки прикладных систем поддержки принятия решений (СППР) и объектно-ориентированных экспертных системах (ЭС). В процесс обучения включаются этапы поиска закономерностей, задания логических правил на базе аппарата формальной или нечеткой логики, составления наборов, проверенных фактов для создания наборов правил соответствия. С точки зрения оценки рисков здоровью населения построение таких моделей МО является довольно трудоемким и затратным процессом в силу длительности процессов получения, валидации, верификации и структурирования данных, а также в поиске экспертизы достаточного уровня надежности [12].

2. Статистическое обучение. При данном подходе процесс построения моделей МО основывается на сборе, обработке, сжатии и преобразовании данных, полученных в результате ведения статистического учета эмпирически полученной информации от специализированных систем, датчиков или операторов, фиксирующих данные в ручном или автоматическом виде. Данный подход является более целесообразным в рамках решаемой нами задачи в силу его большей степени универсальности и возможности автоматизации ряда процессов по предобработке и очистке получаемых данных, что снижает искажения вызванные человеческим фактором [10].

3. Комбинированное обучение, является формой гибридного сочетания статистического и аналитического МО. Это самый сложный из подходов, в силу необходимости интеграции разных подходов и моделей, что не всегда приводит к положительным результатам [13].

В отличие от классических моделей МО, существуют и более сложные модели, применяемые для задач оценок рисков здоровью населения [13]. Отдельного внимания заслуживают подходы, основанные на более сложных моделях МО, базирующихся на принципах построения иерархических связанных структур в виде искусственных нейронных сетей (ИНС).

Следует отметить, что использование различных моделей МО для анализа экологических данных может отличаться математически и алгоритмически в силу специфики их практической и вычислительной имплементации, в том числе для решения задач прогнозирования, выявления аномалий и классификации. В частности, прогнозирование уровней загрязняющих веществ целесообразно осуществлять на базе использования регрессионных моделей, в том числе линейной и полиномиальной регрессий, простых в реализации и достаточно быстрых, моделей градиентного бустинга (XGBoost, LightGBM), позволяющих учесть нелинейные зависимости в данных [14]. Особенности настройки подобных моделей заключаются в выборе регуляризации для уменьшения эффекта переобучения, а также оптимизации подбора значений гиперпараметров, в том числе глубины деревьев и коэффициента обучения. Также возможным является использование методов классификации, в том числе алгоритмов случайного леса, являющегося более устойчивым к шумовым эффектам и эффективным для средних объемов данных, а также метода опорных векторов, который чаще обладает большим уровнем точности при обработке малых наборов данных, для которых явно выражены линейные границы.

Отдельно следует отметить перспективность использования алгоритмов кластеризации, осуществляющих группировку и сегментацию данных по разным критериям, в том числе алгоритмы К-средних и DBSCAN, позволяющих более эффективно и адаптивно отсекаать аномальные значения. При этом для первого из алгоритмов требуется оптимизировать входное число кластеров, например, с помощью применения метода «локтя», а в случае реализации подхода DBSCAN необходимо определить ключевые параметры, такие как минимальный размер кластера и радиус соседства [15].

В силу сложностей обработки больших объемов данных, что является характерным для экологических исследований, еще большую актуальность приобретают модели глубокого обучения, особенно в случаях, когда данные разнородны (текстовые записи с датчиков, графических и табличные данные, спутниковые изображения локаций) и выявление сложных нелинейных зависимостей. В частности, перспективными являются сверточные модели ИНС (для анализа изображений) и рекуррентные (для оценки временных рядов, например, учета динамики загрязнений экосистемы). Сложностями в данном случае является необходимость регулировки внутренней и внешней архитектуры ИНС, в том числе задания значений числа слоев, количества нейронов, а также использование регуляризации [13], [16].

Современные алгоритмы и подходы МО, применяемые для прикладных задач и подходящие для использования в контексте анализа экологических данных, можно классифицировать по четырем подвидам типов обучения [17]:

1. МО с учителем. Данный подход является наиболее популярным и целесообразным для нашей задачи, в связи с тем, что изначально подаваемый набор входных данных проходит разметку в ручном или полу-автоматизированном виде, что позволяет оценить уровень взаимной корреляции. Фактически, выполняется присвоение выходных классов по каждой записи в виде номинальных, категориальных или численных данных, что позволяет применять данный подход для решения задач классификации или регрессии.

2. МО без учителя. Принципиальным отличием от первого типа МО является имплементация процесса обучения без этапа разметки данных. Данный подход обеспечивает возможности просмотра данных и их анализа на основе установки и оценки наиболее значимых связей по степени близости между заданными диапазонами данных. Это позволяет МО без учителя находить слабовыраженные закономерности и корреляцию в признаковом пространстве, что может быть использовано для решения задач кластеризации, поиска аномалий или снижения размерностей, а также совместно с факторным анализом для предобработки данных [11].

3. МО с частичным задействованием учителя. Данный подход является гибридной формой, сочетающей ранее рассмотренные 2 подхода, в котором модели формируются путем использования небольшого объема четко и однозначно размеченных данных (от 10 до 30%), остальные данные первоначально не имеют маркировки. Размеченные записи используются для последующего фрагментарного обучения моделей МО, после чего созданная модель осуществляется разметку имеющихся данных автоматическим образом, фактически, осуществляя псевдомаркировку. На базе данной процедуры реализуется последующее комплексное переобучение модели на полностью размеченной выборке данных [13].

4. Обучение с подкреплением. В данном случае полноценной изначальной разметки данных также не производится, по причине того, что наборы записей для обучения моделей генерируются динамически, поступая из внешней среды, имплементируется механизм автоматического вознаграждения (весовой оценки) моделей МО по совокупности выполненных действий [17]. В силу данной специфики подхода его область применения ограничена средами с активными изменениями, что нехарактерно для оценки рисков здоровья населения.

Несмотря на перспективность и удобство применения моделей МО, следует отметить ряд наиболее принципиальных ограничений их использования непосредственно в экологической сфере, в том числе:

- экологические данные могут содержать значительное количество пропусков, ошибок или шумовых значений, что может быть обусловлено высокой неточностью измерительных приборов или по причине человеческого фактора;
- наличие аномальных значений и выбросов, которые могут быть как реальными данными, так и ошибочными, что существенно усложняет процесс интерпретации результатов;
- недостаток достоверных экологических данных для удаленных или труднодоступных локаций может вносить искажения в точность работы моделей прогнозирования;
- дисбаланс выборок и гетерогенность данных. Следует отметить, что экологические данные являются достаточно разнородными (временные ряды, изображения, пространственные данные), которые не всегда тривиальным образом могут интегрироваться в единую модель, не вызывая при этом противоречий;
- проблемы интерпретации результатов анализа. Многие модели МО, особенно ИИС, не являются четко формализованными с точки зрения внутренних процессов, что в ряде случаев затрудняет интерпретацию их результатов в контексте экологических проблем и задач. В связи с этим данная проблема может стать препятствием для внедрения разработанных и обученных моделей МО в практическую эксплуатацию;
- высокий риск переобучения моделей МО при работе с небольшими объемами данных. В частности, некоторые комплексные модели МО могут слишком сильно адаптироваться к тренировочным данным, уменьшая тем самым их обобщающую способность;
- высокая зависимость работы моделей МО от настройки и подбора значений их гиперпараметров, являющихся в свою очередь трудоемкими процедурами с точки зрения вычислительных ресурсов;
- модели МО, обученные на данных одного региона или экосистемы, могут быть неадаптивны к иным локациям и регионам из-за различий в климатических, географических, биологических или иных условиях, что может потребовать дополнительных затрат на проведения процессов обучения на новых данных.

Заключение

В контексте проанализированной проблематики применение методов машинного обучения является рациональным путем автоматизации процессов анализа экологических данных, в том числе поиска аномалий, оценки корреляции, выявления трендов, заполнения пропусков, балансировки признаков.

С учетом обозначенных ограничений применения машинного обучения в исследуемой сфере предлагается повысить эффективность их имплементации на практике посредством:

- повышения общего уровня качества данных путем их превентивной обработки, в том числе на базе процедур очистки, заполнения пропусков и поиска корреляций для снижения размерности;
- разработка более интерпретируемых (объяснимых) моделей машинного обучения, способных выполнять более четкую формализацию и детализировать логические выводы всех внутренних процессов;
- применение более сложных модификаций алгоритмов, например, ансамблей моделей машинного обучения, что позволит компенсировать отдельные недостатки каждого из используемых алгоритмов;
- применение гибридных подходов к обучению моделей, позволяющих комбинировать модели МО с традиционными аналитико-вероятностными экологическими моделями;
- внедрение практик многоуровневой валидации работы моделей для повышения общего уровня надежности получаемых результатов.

Следует отметить, что в контексте рассматриваемой проблематики перспективным путем дальнейшего развития является разработка и внедрение комплексных подходов к оценке рисков на базе комбинирования методов машинного обучения с различными детерминированными и стохастическими подходами, что позволяет осуществить построение более эффективных гибридных моделей анализа данных, основанных на сочетании как качественных, так и количественных оценок вреда здоровью населения. В частности, среди таких подходов наиболее значимыми являются нейронечеткие модели и алгоритмы, построенные на базе эволюционных вычислений, что может ускорить процессы поиска оптимальных значений гиперпараметров моделей машинного обучения, повышая тем самым адекватность и точность прогнозов.

Конфликт интересов

Не указан.

Conflict of Interest

None declared.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы / References

1. Аскарлов К.А. Оценка риска здоровью населения, проживающего в регионе расположения Жезказганского медеплавильного завода ТОО "KAZAKHMYS SMELTING" / К.А. Аскарлов, К.К. Алимбетов, Г.Т. Байтуякова [и др.] // Вестник Казахского Национального медицинского университета. — 2022. — № 1. — С. 521–529.
2. Имашева Б.С. Оценка риска здоровью населения, проживающего в регионе расположения объектов Павлодарского алюминиевого завода АО «Алюминий Казахстана» / Б.С. Имашева, К.А. Аскарлов, М.С. Имашев [и др.] // Наука и Здравоохранение. — 2022. — № 3(24). — С. 66–77.
3. Камдина Л.В. Развитие методического инструментария оценки влияния антропогенных факторов промышленного производства на качество жизни населения : автореф. дис. ... канд. экон. наук / Л.В. Камдина. — 2019. — 23 с.
4. Rong G. Population amount risk assessment of extreme precipitation-induced landslides based on integrated machine learning model and scenario simulation / G. Rong, T. Kaiwei, L. Zhijun [et al.] // Geoscience Frontiers. — 2023. — № 14. DOI: 10.1016/j.gsf.2023.101541
5. Alizadeh M. Human health risk of nitrate in groundwater of Tehran–Karaj plain, Iran / M. Alizadeh, R. Noori, B. Omidvar [et al.] // Scientific Reports. — 2024. — № 14. DOI: 10.1038/s41598-024-58290-6
6. Ermolaeva S. Evaluation of environmental and public health status under anthropogenic disturbance / S. Ermolaeva, D. Khisamutdinov // Ulyanovsk Medico-biological Journal. — 2023. — P. 145–154. DOI: 10.34014/2227-1848-2023-2-145-154
7. Виноградова Е.А. Способы экологического мониторинга воздуха с применением технологических решений для их обработки и анализа, включая машинное обучение, искусственный интеллект и аналитику больших данных / Е.А. Виноградова, М.М. Дмитриев, А.С. Дудрявец [и др.] // Современные технологии: проблемы и тенденции развития. — Петрозаводск : Международный центр научного партнерства «Новая Наука», 2021. — С. 174–189.
8. Железный С.В. Использование методов машинного обучения для прогнозирования загрязненности атмосферного воздуха / С.В. Железный, А.И. Ситников, А.А. Толстых // Вестник Воронежского института МВД России. — 2019. — № 3. — С. 73–81.
9. Тагиева Я.Р. Использование методов машинного обучения для прогнозирования уровня загрязнения воздуха / Я.Р. Тагиева // Устойчивое развитие науки и образования. — 2020. — № 3(42). — С. 93–97.
10. Bataresh F.A. Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering / F.A. Bataresh. — London : Academic Press, 2020. — 266 p.
11. Лохов А.С. Сравнительный анализ применения весовых коэффициентов и коэффициентов значимости в классификационных геоэкологических моделях / А.С. Лохов, В.Б. Коробов // Проблемы региональной экологии. — 2022. — № 4. — С. 81–86.

Список литературы на английском языке / References in English

1. Askarov K.A. Ocenka riska zdorov'ju naselenija, prozhivajushhego v regione raspolozhenija Zhezkazganskogo medeplavil'nogo zavoda TOO "KAZAKHMYS SMELTING" [Assessment of the health risk of the population living in the region of the Zhezkazgan copper smelter of KAZAKHMYS SMELTING LLP] / K.A. Askarov, K.K. Alimbetov, G.T. Baytuyakova [et al.] // Vestnik Kazahskogo Nacional'nogo medicinskogo universiteta [Bulletin of the Kazakh National Medical University]. — 2022. — № 1. — P. 521–529. [in Russian]
2. Imasheva B.S. Ocenka riska zdorov'ju naselenija, prozhivajushhego v regione raspolozhenija ob#ektov Pavlodarskogo aljuminievogo zavoda AO «Aljuminij Kazahstana» [Assessment of the health risk of the population living in the region of the location of the Pavlodar aluminum plant of JSC Aluminum of Kazakhstan] / B.S. Imasheva, K.A. Askarov, M.S. Imashev [et al.] // Nauka i Zdravoohranenie [Science and Healthcare]. — 2022. — № 3(24). — P. 66–77. [in Russian]
3. Kamdina L.V. Razvitie metodicheskogo instrumentarija ocenki vlijanija antropogennyh faktorov promyshlennogo proizvodstva na kachestvo zhizni naselenija [Development of methodological tools for assessing the impact of anthropogenic factors of industrial production on the quality of life of the population] : abstract of the dissertation. ... Candidate of Economic Sciences / L.V. Kamdina. — 2019. — 23 p. [in Russian]
4. Rong G. Population amount risk assessment of extreme precipitation-induced landslides based on integrated machine learning model and scenario simulation / G. Rong, T. Kaiwei, L. Zhijun [et al.] // Geoscience Frontiers. — 2023. — № 14. DOI: 10.1016/j.gsf.2023.101541
5. Alizadeh M. Human health risk of nitrate in groundwater of Tehran–Karaj plain, Iran / M. Alizadeh, R. Noori, B. Omidvar [et al.] // Scientific Reports. — 2024. — № 14. DOI: 10.1038/s41598-024-58290-6
6. Ermolaeva S. Evaluation of environmental and public health status under anthropogenic disturbance / S. Ermolaeva, D. Khisamutdinov // Ulyanovsk Medico-biological Journal. — 2023. — P. 145–154. DOI: 10.34014/2227-1848-2023-2-145-154
7. Vinogradova E.A. Sposoby jekologicheskogo monitoringa vozduha s primeneniem tehnologicheskikh reshenij dlja ih obrabotki i analiza, vkljuchaja mashinnoe obuchenie, iskusstvennyj intellekt i analitiku bol'shih dannyh [Methods of environmental air monitoring using technological solutions for their processing and analysis, including machine learning, artificial intelligence and big data analytics] / E.A. Vinogradova, M.M. Dmitriev, A.S. Dudryavets [et al.] // Sovremennye tehnologii: problemy i tendencii razvitija [Modern Technologies: Problems and Development Trends]. — Petrozavodsk : International Center for Scientific Partnership "New Science", 2021. — P. 174–189. [in Russian]
8. Zhelezny S.V. Ispol'zovanie metodov mashinnogo obuchenija dlja prognozirovanija zagrjaznennosti atmosfernogo vozduha [The use of machine learning methods for predicting atmospheric air pollution] / S.V. Zhelezny, A.I. Sitnikov, A.A. Tolstykh // Vestnik Voronezhskogo instituta MVD Rossii [Bulletin of the Voronezh Institute of the Ministry of Internal Affairs of Russia]. — 2019. — № 3. — P. 73–81. [in Russian]
9. Tazieva Ya.R. Ispol'zovanie metodov mashinnogo obuchenija dlja prognozirovanija urovnja zagrjaznenija vozduha [Using machine learning methods to predict air pollution levels] / Ya.R. Tazieva // Ustojchivoje razvitie nauki i obrazovanija [Sustainable Development of Science and Education]. — 2020. — № 3(42). — P. 93–97. [in Russian]
10. Bataresh F.A. Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering / F.A. Bataresh. — London : Academic Press, 2020. — 266 p.
11. Lokhov A.S. Sravnitel'nyj analiz primenenija vesovyh koeficientov i koeficientov znachimosti v klassifikacionnyh geojekologicheskikh modeljah [Comparative analysis of the application of weighting coefficients and significance coefficients in classification geoeological models] / A.S. Lokhov, V.B. Korobov // Problemy regional'noj jekologii [Problems of Regional Ecology]. — 2022. — № 4. — P. 81–86. [in Russian]