

АВТОМАТИЗАЦИЯ СИСТЕМЫ УПРАВЛЕНИЯ ПРОЦЕССАМИ ОЧИСТКИ СТОЧНЫХ ВОД С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Игорь Алексеевич Гульшин

Национальный исследовательский Московский государственный строительный университет
(НИУ МГСУ); г. Москва, Российская Федерация

В статье рассматривается алгоритм сбора и анализа технологических данных очистных сооружений с использованием методов машинного обучения. Основная цель исследования заключается в решении двух ключевых задач: регрессионного прогнозирования качества очищенных сточных вод и классификации с целью предотвращения аварийных ситуаций на очистных сооружениях, в частности, нитчатого вспухания активного ила. Поскольку получение больших объемов качественных данных непосредственно на действующих очистных установках часто затруднено, особое внимание уделено оценке возможности использования лабораторно полученных данных в качестве обучающего набора. Для этого был сформирован датасет в условиях, воспроизводящих реальный технологический процесс, с последующим применением его для обучения моделей. Тестирование и валидация алгоритма проводились на небольшом наборе реальных производственных данных, полученных непосредственно с действующих станций очистки сточных вод. В результате экспериментального сравнения нескольких подходов к решению указанных задач было выявлено, что наилучшие показатели точности демонстрируют модели градиентного бустинга из семейства CatBoost. Для задачи регрессии удалось достичь среднего абсолютного процентного отклонения (SMAPE) на уровне 9,1 %, а при решении задачи бинарной классификации была получена максимально возможная метрика ROC-AUC, равная 1,0. Кроме того, в рамках работы определен набор наиболее значимых предикторов для каждой из прогнозируемых целевых переменных, что позволило выявить ключевые факторы, влияющие на эффективность очистки и стабильность работы очистных сооружений. Предложены практические рекомендации по подготовке данных и выбору оптимальных моделей, что подтверждает перспективность подхода для автоматизированного контроля и интеллектуального управления процессами очистки сточных вод.

Ключевые слова: очистка сточных вод, машинное обучение, градиентный бустинг, CatBoost, прогнозирование, нитчатое вспухание, подготовка данных

Для цитирования: Гульшин И.А. Автоматизация системы управления процессами очистки сточных вод с помощью методов машинного обучения // Биосферная совместимость: человек, регион, технологии. 2025. № 1. С. 96–103. DOI: 10.22227/2311-1518.2025.1.96-103

AUTOMATION OF WASTEWATER TREATMENT PROCESS CONTROL SYSTEM USING MACHINE LEARNING METHODS

Igor A. Gulshin

Moscow State University of Civil Engineering (National Research University) (MGSU);
Moscow, Russian Federation

The paper examines an algorithm for the collection and analysis of technological data from wastewater treatment facilities using machine learning methods. The primary objective of this research is to solve two key tasks: regression-based forecasting of treated wastewater quality and classification to prevent emergency situations at treatment facilities, particularly filamentous bulking of activated sludge. Since obtaining large volumes of high-quality data directly from operational wastewater treatment plants is often challenging, special attention is given to evaluating the feasibility of using laboratory-generated data as a training set. A dataset replicating real technological processes was created in laboratory conditions and subsequently used for model training. The algorithm was tested and validated using a small set of actual operational data obtained directly from existing wastewater treatment plants. Through experimental comparison of various approaches for solving the defined tasks, gradient boosting models from the CatBoost family demonstrated the best accuracy metrics. In the regression task, a Symmetric Mean Absolute Percentage Error (SMAPE) of 9.1 % was achieved, while the binary classification task yielded the maximum possible ROC-AUC metric of 1.0. Furthermore, the study identified a set of the most significant predictors for each target variable, highlighting key factors affecting treatment efficiency and operational stability. Practical recommendations on data preparation and optimal model selection are proposed, demonstrating the potential of the developed approach for automated monitoring and intelligent management of wastewater treatment processes.

Keywords: wastewater treatment, machine learning, gradient boosting, CatBoost, forecasting, filamentous bulking, data preparation

For citation: Gulshin I.A. Automation of wastewater treatment process control system using machine learning methods. *Biosphere Compatibility: Man, Region, Technology*. 2025; 1:96-103. DOI: 10.22227/2311-1518.2025.1.96-103 (rus.).

Введение

Очистка сточных вод является критически важной деятельностью, непосредственно влияющей на устойчивое развитие комфортной среды проживания. Она необходима во всех типах производственных процессов и выступает ключевым элементом инфраструктуры населенных пунктов любого масштаба. Несмотря на значительные капитальные вложения, не всегда удается стабильно повысить качество очистки сточных вод. Современные исследования работы существующих очистных сооружений показывают, что неудовлетворительные результаты зачастую обусловлены эксплуатационными ошибками. Хотя современные станции очистки оснащены автоматизированными системами управления, они эффективны лишь при штатном режиме работы. В реальности биологические очистные сооружения представляют собой высокочувствительные биореакторы, функционирующие в квазистационарных условиях. Даже незначительные отклонения в технологических параметрах могут привести к аварийным ситуациям, снижению качества очистки и технологическим сбоям. В условиях нехватки квалифицированного персонала оперативное реагирование на возникающие проблемы затруднено. Это ведет к рискам технологических, экологических и санитарных происшествий, сопровождаемых значительными финансовыми потерями.

В последние годы существенно возрос интерес к применению методов машинного обучения для прогнозирования работы очистных сооружений. Исследования направлены на предсказание качества очищенных сточных вод и своевременное обнаружение аварийных ситуаций [1–7]. Наибольшее внимание уделяется двум типам задач: регрессионному прогнозированию и классификации аномалий. Регрессионные модели ориентированы на внедрение «виртуальных сенсоров», призванных снизить эксплуатационные затраты на измерительные приборы. Например, Zhang [1] применил модель XGBoost для прогнозирования эффективности удаления азота и фосфора с использованием данных о видовом составе активного ила. Модель показала коэффициент детерминации более 0,8. Однако использование таких специфических признаков, как 16S рРНК, затруднено на большинстве реальных очистных сооружений из-за технической сложности анализа.

В работе Wang [2] рассмотрено влияние эксплуатационных параметров на очистной станции в г. Умео (Швеция). На основе анализа важности признаков (Variable Importance Measure, VIM) были сформированы гипотезы о технологических взаимо-

связях на объекте. Ограничением модели стала ее работа по принципу «черного ящика», однако авторы подчеркнули важность учета временных лагов. Проблему учета временных лагов также исследовал Ху [3], который предложил модель на основе нейронной сети LSTM для прогноза качества очистки сточных вод с учетом задержки обработки сточных вод. Средняя абсолютная процентная ошибка (MAPE) составила 2,3 %, значительно ниже по сравнению с традиционными методами. Однако важно отметить, что производительность таких моделей сильно зависит от качества используемых данных [4]. Модели LSTM склонны к переобучению на малых наборах данных и чувствительны к выбросам [5]. Тем не менее ряд исследователей успешно применяют подходы временных рядов для прогнозирования качества очистки [6, 7]. В качестве альтернативы был предложен подход ORELM (Outlier Robust Extreme Learning Machine), модифицированный специально для работы с выбросами, которые характерны для технологических данных очистных сооружений. Этот метод показал высокие результаты с MAPE около 1,0 %, подтверждая перспективность использования специализированных методов обработки выбросов [8–11].

Задачи классификации в основном ориентированы на выявление аномалий. Bellamoli [9] разработал подход классификации аномалий для реакторов периодического действия (SBR) на основе моделей ансамблей деревьев решений. Максимальный recall достиг 0,83. Elsayed [10] и Nasir [11] применили различные модели классификации, в том числе CatBoost, получивший точность классификации 94,51 %. CatBoost выгодно отличается эффективной обработкой категориальных признаков и высокой скоростью обучения [12–15].

Для решения перечисленных проблем настоящая работа предлагает стратегию формирования обучающих наборов данных в лабораторных условиях с последующей валидацией на реальных данных очистных сооружений. В ходе исследования была проверена гипотеза о применимости лабораторных данных, разработаны рекомендации по подготовке данных, выбору признаков и моделей. Особое внимание уделено использованию признаков, доступных на большинстве объектов, таких как концентрация кислорода и температура. По итогам экспериментальной части работы были разработаны рекомендации по обработке данных, включающие учет соотношения углерода к азоту (C/N), исключение полных дубликатов, обработку пропусков методом Pairwise Deletion и использование робаст-

ных методов масштабирования. Оптимальными моделями были признаны градиентные бустинги семейства CatBoost с обязательной процедурой кросс-валидации для настройки гиперпараметров. Качество моделей оценивалось с помощью метрик SMAPE и ROC-AUC, что позволило достичь высокой точности прогнозирования. Разработанный подход подтверждает свою эффективность для интеллектуального управления процессами очистки сточных вод и предотвращения аварийных ситуаций.

Материалы и методы

Основная задача исследования состояла в разработке метода анализа. Основной целью исследования являлась оценка возможности использования обучающих и тестовых наборов данных из различных источников для прогнозирования качества очищенных сточных вод и выявления аномалий технологического процесса. Тестовый датасет состоял из результатов технологического мониторинга очистных сооружений Московской области, проведенного в 2022–2023 гг., и включал 140 записей по 15 параметрам. Ввиду ограниченности данных для формирования полноценного обучающего набора был задействован лабораторный биореактор, позволяющий генерировать распределения параметров, аналогичные реальным данным.

В ходе лабораторного эксперимента получен набор из 4303 записей по 15 параметрам, характеризующим качество поступающих и очищенных сточных вод (биологическая стадия очистки). Определялись биохимическое потребление кислорода (БПК₅), химическое потребление кислорода (ХПК), аммонийный азот (NH₄), ортофосфаты (PO₄), нитраты и нитриты (NO₂, NO₃ — только в очищенных водах). Технологические параметры включали температуру сточных вод, концентрацию растворенного кислорода в аэробной зоне и долю углекислого газа в отводимом воздухе. Категориальные признаки включали тип внешнего источника углерода (ацетат или раствор синтезированных летучих жирных кислот после подкисления), точку подачи внешнего источника и наличие нитчатого вспухания активного ила. Средние значения параметров представлены в таблице.

При предварительной обработке данных были осуществлены проверка типов данных, выявление и обработка пропусков и дубликатов. Пропуски обрабатывались методом Pairwise Deletion, замещаясь заполнителями, которые могли быть временно исключены из анализа при помощи логической фильтрации. Для временных рядов применялась автоматизированная импутация. Статистический анализ включал проверку нормальности распределений по критерию Шапиро–Уилка. В случае

Результаты анализа основных нормативно-технических документов в части требований к оценке карстоопасности территорий

Параметр	Лабораторная установка	Очистные сооружения
БПК ₅ , мгО ₂ /л	107	108
ХПК, мгО ₂ /л	135	136
Аммонийный азот, мг/л	35	34,6
Ортофосфаты, мг/л	8	7,6
Растворенный кислород, мг/л	2,51	2,53
Температура, °C	20	19,8

ненормальности распределений использовался Bootstrap-анализ с выборками в 10 % от общего размера и оценкой значимости при уровне 0,05. При подтверждении равенства средних значений выборок проводился корреляционный анализ с использованием инструмента `phik_matrix` библиотеки `Phik`, подходящего для оценки нелинейных и категориальных зависимостей.

Эксперимент проводился с использованием автоматизированного лабораторного биореактора-ферментера Yocell YC-JG (Yocell Biotechnology, Китай) с рабочим объемом 11 литров. Фотография биореактора представлена на рисунке. Реактор оснащен контролируемой электромеханической мешалкой, пневматической системой аэрации и системой термостатирования. Аналитическое оборудование включало сенсоры растворенного кислорода, pH, соединений азота, концентрации углекислого газа, а также приборы для анализа ХПК и БПК (Hamilton, HACH, BlueSens, Xylem Analytics). Контрольные измерения проводились спектрофотометром HACH Lange DR6000 и другими стандартными аналитическими методами. Управление осуществлялось с помощью программного обеспечения Siemens Simatic. В качестве субстрата использовалась сточная вода и растворы летучих жирных кислот, полученных ферментацией осадка.

В работе решались задачи регрессии (прогнозирование численных характеристик очищенных сточных вод) и классификации (выявление нитчатого вспухания активного ила). Процесс моделирования включал обработку данных, рандомизированную кросс-валидацию моделей и оценку значимости признаков методом SHAP. Для регрессии использовались модели DecisionTreeRegressor, LinearRegression, ElasticNet, Lasso Regression, Ridge Regression, XGBoost, LightGBM и CatBoost с метрикой SMAPE. Качество моделей дополнительно оценивалось метриками R², RMSE и MAE.

Для классификации применялись модели DecisionTreeClassifier, LogisticRegression, KNeighborsClassifier, SVC, а также XGBoost, LightGBM и



Лабораторная установка, используемая для набора тренировочной выборки

CatBoost с метрикой ROC-AUC. Настройка гиперпараметров осуществлялась при помощи RandomizedSearchCV. Проверка адекватности моделей проводилась сравнением с постоянной моделью DummyRegressor. Тестовая выборка формировалась из реальных данных очистных сооружений после предварительного статистического анализа.

Результаты и обсуждение

Перед построением моделей был проведен комплексный статистический анализ наборов данных, полученных с лабораторной установки и действующих очистных сооружений. Установлено, что численные признаки имеют разные шкалы измерений, что потребовало масштабирования данных перед моделированием. Анализ также выявил отсутствие аномальных значений, подтверждая пригодность данных для дальнейшей работы. Сравнительное исследование подтвердило соответствие природы признаков в обоих наборах данных.

Для проверки гипотезы о равенстве соотношений признаков между лабораторными и производственными выборками был введен дополнительный признак — соотношение легкоокисляемой органики (БПК₅) к аммонийному азоту (NH₄), а также показатели эффективности очистки по основ-

ным загрязнениям. Распределения численных признаков были близки к нормальному, хотя были выявлены выбросы, соответствующие естественной изменчивости процесса очистки сточных вод.

Для устранения негативного влияния дисбаланса классов при моделировании аварийной ситуации (нитчатое вспухание активного ила) была применена процедура oversampling. Результаты анализа Bootstrap подтвердили равенство средних значений выборок по основным показателям, что позволило обоснованно использовать лабораторные данные для моделирования реальных условий.

Корреляционный анализ не выявил значимой мультиколлинеарности между входными признаками. При этом отмечена сильная корреляция между отдельными целевыми параметрами, такими как ХПК/БПК₅, точка подачи субстрата/фосфаты, нитраты/нитриты. Это связано с биохимической природой процессов, где реакционная способность существенно зависит от температуры и концентрации растворенного кислорода. В результате применения процедуры перекрестной валидации и анализа важности признаков (SHAP-анализ) лучшими моделями были признаны алгоритмы градиентного бустинга семейства CatBoost. Для регрессионных задач достигнуты значения SMAPE менее 10 %,

а для классификации аварийных ситуаций метрика ROC-AUC составила 0,95.

При решении задач классификации особое внимание уделялось исключению ложноотрицательных прогнозов, что было подтверждено матрицей ошибок, свидетельствующей об отсутствии необходимости корректировки пороговых значений модели. Анализ остатков для регрессионных задач не выявил систематической ошибки модели, однако указал на необходимость расширения набора данных для повышения точности прогнозирования. В целом полученные результаты демонстрируют перспективность применения моделей CatBoost для прогнозирования качества очищенных сточных вод и предотвращения аварийных ситуаций на очистных сооружениях. Включение специфических технологических параметров (например, типа и точки ввода внешнего субстрата) позволило повысить точность и интерпретируемость моделей, подтвердив важность комплексного подхода к анализу данных в задачах управления очисткой сточных вод.

Нитчатое вспухание активного ила классифицируется как аварийная ситуация в системах очистки сточных вод, приводящая к значительным негативным последствиям для технологического процесса. В частности, нитчатое вспухание часто становится причиной полного нарушения работы системы из-за потери осаждаемости ила, что ведет к резкому ухудшению качества очищенных сточных вод. Вероятность возникновения такой ситуации при обычных условиях эксплуатации оценивается как низкая (около 3 % по данным лабораторного эксперимента). В связи с этим при обучении моделей важна стратификация данных или применение метода oversampling для устранения негативного влияния дисбаланса классов. Использование данных методов гарантирует, что редкие, но критические события будут адекватно представлены в обучающем наборе, улучшая прогнозную способность модели.

Анализ распределений признаков показал, что большинство из них близки к нормальному, однако выявлено отклонение от нормальности, обусловленное пределами обнаружения аналитического оборудования. Данное обстоятельство усложняет проведение статистического анализа, предполагающего нормальность распределений, и требует использования робастных статистических методов, таких как Bootstrap-анализ. Проведенный Bootstrap-анализ подтвердил равенство средних значений показателей лабораторных и эксплуатационных выборок, что обосновывает возможность применения лабораторных данных для моделирования реальных процессов очистки сточных вод. Построенные диаграммы рассеяния подтвердили наличие линейной зависимости только для аммонийного азота, раство-

ренного кислорода и температуры. Это обусловлено природой биохимических процессов, интенсивность которых зависит от температуры среды и концентрации кислорода, доступного для микроорганизмов. Значимую роль также играет доступность органического субстрата относительно количества поступающего азота (соотношение C/N).

Отсутствие явных линейных зависимостей между большинством предикторов и целевыми переменными свидетельствует о необходимости применения нелинейных методов машинного обучения, таких как градиентный бустинг. В результате исследования наиболее точные прогнозы были получены с использованием модели CatBoost, обладающей высокой эффективностью обработки категориальных признаков. Анализ важности признаков (SHAP) позволил выявить значимость ряда специфических технологических параметров, таких как тип и точка ввода внешнего субстрата, концентрация растворенного кислорода и температура. Эти результаты существенно дополняют понимание факторов, влияющих на качество очистки.

Особое внимание в классификационных задачах уделялось минимизации ложноотрицательных прогнозов (пропусков аварийных ситуаций). Матрица ошибок показала, что модель успешно идентифицировала все случаи нитчатого вспухания, исключив необходимость изменения пороговых значений модели. Проведенный анализ остатков регрессионных моделей подтвердил отсутствие систематических отклонений, однако выявил необходимость увеличения объема исторических данных для улучшения качества прогнозов.

В целом исследование подтверждает целесообразность использования методов машинного обучения, в частности модели CatBoost, для прогнозирования качества очищенных сточных вод и своевременного выявления аварийных ситуаций. Дальнейшие исследования должны быть направлены на изучение влияния типа и точек введения внешних субстратов на эффективность очистки, а также на расширение и диверсификацию используемых наборов данных.

Выводы

В рамках выполненного исследования основной задачей являлась не разработка конкретных моделей, а формулирование подходов и практических рекомендаций по сбору данных, подготовке их для моделирования и последующему выбору оптимальных моделей. В связи с высокой чувствительностью моделей к изменениям технологических параметров (например, продолжительности очистки) любые корректировки требуют обязательного переобуче-

ния модели. На основе проведенной работы сформулированы следующие рекомендации:

1. Необходимо строго соблюдать соответствие лабораторного моделирования реальным технологическим параметрам исследуемых очистных сооружений. В первую очередь это касается средних концентраций входящих загрязняющих веществ, растворенного кислорода и температуры. Доля углекислого газа, как нерегулируемый показатель, может быть исключена при значительных расхождениях. Желательно использовать аналитическое оборудование, аналогичное установленному на реальных объектах, а также автоматизировать сбор показаний датчиков, ориентируясь на максимально возможные интервалы измерений.

2. Из обучающего набора следует удалять полные дубликаты данных, предотвращая тем самым утечку информации при кросс-валидации. Для количественных признаков рекомендуется использовать стратегию Pairwise Deletion. Пропущенные значения временных рядов следует заполнять средним значением двух соседних точек, а для категориальных признаков — модой.

3. Не рекомендуется удалять выбросы, находящиеся за пределами межквартильного размаха, так как они часто сигнализируют об аварийных ситуациях. Перед обучением модели необходимо проводить статистические тесты на равенство средних значений выборки (например, критерий Стьюдента при выборках более 50 наблюдений). Также обязателен корреляционный анализ для выявления мультиколлинеарности. Если данные не подходят для стандартных статистических тестов, рекомендуется Bootstrap-анализ, так как показатели эффективности очистки воды часто имеют экспоненциальное распределение.

4. При значительном дисбалансе классов целесообразно применять методы oversampling или стратификацию данных. В настоящем исследовании высокое качество предсказаний было достигнуто

и при дисбалансе классов, однако необходимость использования данных методов должна проверяться отдельно для каждого конкретного случая. Масштабирование признаков рекомендуется проводить робастными методами, такими как RobustScaler.

5. Для рассматриваемого набора признаков наиболее эффективными оказались модели градиентного бустинга семейства CatBoost, применимые как для регрессионных, так и для классификационных задач. Поскольку для каждой целевой переменной существуют собственные оптимальные гиперпараметры, следует использовать процедуру кросс-валидации для их подбора.

6. В качестве наиболее объективных метрик оценки моделей предлагается использовать SMAPE для регрессии и ROC-AUC для классификации. Адекватность моделей рекомендуется проверять путем сравнения полученных метрик с метриками постоянной модели.

7. При решении классификационных задач по выявлению аварийных ситуаций необходимо минимизировать вероятность ложноотрицательных прогнозов, что означает исключение пропуска предсказаний аварий, кодируемых как «0» при помощи LabelEncoder. Для этого требуется проводить оценку качества классификации по confusion matrix и, при необходимости, повторно настраивать пороговые значения моделей по метрикам Recall и Precision.

8. Полученные модели позволяют прогнозировать как эффективность очистки сточных вод, так и вероятность возникновения аварийных ситуаций с учетом временного лага, равного гидравлическому времени пребывания (HRT). Величина HRT закладывается в проектные расчеты очистных сооружений и корректируется в зависимости от гидравлической нагрузки на биореактор. Этот фактор должен учитываться при использовании моделей в каждом конкретном случае.

СПИСОК ИСТОЧНИКОВ

1. Zhang Y., Wu H., Xu R., Wang Y., Chen L., Wei C. Machine learning modeling for the prediction of phosphorus and nitrogen removal efficiency and screening of crucial microorganisms in wastewater treatment plants // *Sci. Total Environ.* 2024. Vol. 907. No. 167730. DOI: 10.1016/j.scitotenv.2023.167730
2. Wang D., Thunéll S., Lindberg U., Jiang L., Trygg J., Tysklind M. et al. A machine learning framework to improve effluent quality control in wastewater treatment plants // *Sci. Total Environ.* 2021. Vol. 784. No. 147138. DOI: 10.1016/j.scitotenv.2021.147138
3. Xu B., Pooi C.K., Tan K.M., Huang S., Shi X., Ng H.Y. A novel long short-term memory artificial neural network (LSTM)-based soft-sensor to monitor and forecast wastewater treatment performance // *J. Water Process Eng.* 2023. Vol. 54. No. 104041. DOI: 10.1016/j.jwpe.2023.104041
4. Abouzari M., Pahlavani P., Izaditame F., Bigdeli B. Estimating the chemical oxygen demand of petrochemical wastewater treatment plants using linear and nonlinear statistical models : a case study // *Chemosphere.* 2021. Vol. 270. No. 129465. DOI: 10.1016/j.chemosphere.2020.129465
5. Recio-Colmenares R., León Becerril E., Gurubel Tun K.J., Conchas R.F. Design of a Soft Sensor Based on

- Long Short-Term Memory Artificial Neural Network (LSTM) for Wastewater Treatment Plants // *Sensors*. 2023. Vol. 23. Issue 22. No. 9236. DOI: 10.3390/s23229236
6. El-Rawy M., Abd-Ellah M.K., Fathi H., Ahmed A.K.A. Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques // *J. Water Process Eng.* 2021. Vol. 44. No. 102380. DOI: 10.1016/j.jwpe.2021.102380
 7. Singh N.K., Yadav M., Singh V., Padhiyar H., Kumar V., Bhatia S.K. et al. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems // *Bioresour. Technol.* 2023. Vol. 369. No. 128486. DOI: 10.1016/j.biortech.2022.128486
 8. Boumezbeur H., Laouacheria F., Heddam S., Djemili L. Modelling coagulant dosage in drinking water treatment plant using advance machine learning model: Hybrid extreme learning machine optimized by Bat algorithm // *Environ. Sci. Pollut. Res.* 2023. Vol. 30. No. 28. Pp. 72463–72483. DOI: 10.1007/s11356-023-27224-6
 9. Bellamoli F., Di Iorio M., Vian M., Melgani F. Machine learning methods for anomaly classification in wastewater treatment plants // *J. Environ. Manage.* 2023. Vol. 344. No. 118594. DOI: 10.1016/j.jenvman.2023.11859411
 10. Elsayed A., Siam A., El-Dakhakhni W. Machine learning classification algorithms for inadequate wastewater treatment risk mitigation // *Process Saf Environ Prot.* 2022. Vol. 159. Pp. 1224–1235. DOI: 10.1016/j.psep.2022.01.06512
 11. Nasir N., Kansal A., Alshaltone O., Barneih F., Sameer M., Shanableh A. et al. Water quality classification using machine learning algorithms // *J. Water Process Eng.* 2022. Vol. 48. No. 102920. DOI: 10.1016/j.jwpe.2022.10292013
 12. Jiang J., Xiang X., Zhou Q., Zhou L., Bi X., Khanal S.K. et al. Optimization of a Novel Engineered Ecosystem Integrating Carbon, Nitrogen, Phosphorus, and Sulfur Biotransformation for Saline Wastewater Treatment Using an Interpretable Machine Learning Approach // *Environ. Sci. Technol.* 2024. Vol. 58. No. 29. Pp. 12989–12999. DOI: 10.1021/acs.est.4c0316014
 13. Al Nuaimi H., Abdelmagid M., Bouabid A., Chrysikopoulos C.V., Maalouf M. Classification of WatSan Technologies using machine learning techniques // *Water*. 2023. Vol. 15. Issue 15. No. 2829. DOI: 10.3390/w1515282915
 14. Wang Q., Li Z., Cai J., Zhang M., Liu Z., Xu Y., Li R. Spatially adaptive machine learning models for predicting water quality in Hong Kong // *J. Hydrol.* 2023. Vol. 622. No. 129649. DOI: 10.1016/j.jhydrol.2023.12964916
 15. Halalsheh N., Alshboul O., Shehadeh A., Al Mamlook R.E., Al-Othman A., Tawalbeh M. et al. Breakthrough curves prediction of selenite adsorption on chemically modified zeolite using boosted decision tree algorithms for water treatment applications // *Water*. 2022. Vol. 14. Issue 16. No. 2519. DOI: 10.3390/w14162519

Об авторе: **Игорь Алексеевич Гульшин** — доцент кафедры водоснабжения и водоотведения; **Национальный исследовательский Московский государственный строительный университет (НИУ МГСУ)**; 129337, г. Москва, Ярославское шоссе, д. 26; e-mail: GulshinIA@mgsu.ru.

REFERENCES

1. Zhang Y., Wu H., Xu R., Wang Y., Chen L., Wei C. Machine learning modeling for the prediction of phosphorus and nitrogen removal efficiency and screening of crucial microorganisms in wastewater treatment plants. *Sci. Total Environ.* 2024; 907:167730. DOI: 10.1016/j.scitotenv.2023.167730
2. Wang D., Thunell S., Lindberg U., Jiang L., Trygg J., Tysklind M. et al. A machine learning framework to improve effluent quality control in wastewater treatment plants. *Sci. Total Environ.* 2021; 784:147138. DOI: 10.1016/j.scitotenv.2021.147138
3. Xu B., Pooi C.K., Tan K.M., Huang S., Shi X., Ng H.Y. A novel long short-term memory artificial neural network (LSTM)-based soft-sensor to monitor and forecast wastewater treatment performance. *J. Water Process Eng.* 2023; 54:104041. DOI: 10.1016/j.jwpe.2023.104041
4. Abouzari M., Pahlavani P., Izaditame F., Bigdeli B. Estimating the chemical oxygen demand of petrochemical wastewater treatment plants using linear and nonlinear statistical models : A case study. *Chemosphere.* 2021; 270:129465. DOI: 10.1016/j.chemosphere.2020.129465
5. Recio-Colmenares R., León Becerril E., Gurubel Tun K.J., Conchas R.F. Design of a Soft Sensor Based on Long Short-Term Memory Artificial Neural Network (LSTM) for Wastewater Treatment Plants. *Sensors.* 2023; 23(22):9236. DOI: 10.3390/s23229236
6. El-Rawy M., Abd-Ellah M.K., Fathi H., Ahmed A.K.A. Forecasting effluent and performance of wastewater treatment plant using different machine learning techniques. *J. Water Process Eng.* 2021; 44:102380. DOI: 10.1016/j.jwpe.2021.102380

7. Singh N.K., Yadav M., Singh V., Padhiyar H., Kumar V., Bhatia S.K. et al. Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems. *Bioresour. Technol.* 2023; 369:128486. DOI: 10.1016/j.biortech.2022.128486
8. Boumezbeur H., Laouacheria F., Heddam S., Djemili L. Modelling coagulant dosage in drinking water treatment plant using advance machine learning model: Hybrid extreme learning machine optimized by Bat algorithm. *Environ. Sci. Pollut. Res.* 2023; 30(28):72463-72483. DOI: 10.1007/s11356-023-27224-6
9. Bellamoli F., Di Iorio M., Vian M., Melgani F. Machine learning methods for anomaly classification in wastewater treatment plants. *J. Environ. Manage.* 2023; 344:118594. DOI: 10.1016/j.jenvman.2023.11859411
10. Elsayed A., Siam A., El-Dakhkhni W. Machine learning classification algorithms for inadequate wastewater treatment risk mitigation. *Process Saf Environ Prot.* 2022; 159:1224-1235. DOI: 10.1016/j.psep.2022.01.06512
11. Nasir N., Kansal A., Alshaltone O., Barneih F., Sameer M., Shanableh A. et al. Water quality classification using machine learning algorithms. *J. Water Process Eng.* 2022; 48:102920. DOI: 10.1016/j.jwpe.2022.10292013
12. Jiang J., Xiang X., Zhou Q., Zhou L., Bi X., Khanal S.K. et al. Optimization of a Novel Engineered Ecosystem Integrating Carbon, Nitrogen, Phosphorus, and Sulfur Biotransformation for Saline Wastewater Treatment Using an Interpretable Machine Learning Approach. *Environ. Sci. Technol.* 2024; 58(29):12989-12999. DOI: 10.1021/acs.est.4c0316014
13. Al Nuaimi H., Abdelmagid M., Bouabid A., Chrysikopoulos C.V., Maalouf M. Classification of WatSan Technologies using machine learning techniques. *Water.* 2023; 15(15):2829. DOI: 10.3390/w1515282915
14. Wang Q., Li Z., Cai J., Zhang M., Liu Z., Xu Y. et al. Spatially adaptive machine learning models for predicting water quality in Hong Kong. *J. Hydrol.* 2023; 622:129649. DOI: 10.1016/j.jhydrol.2023.12964916
15. Halalsheh N., Alshboul O., Shehadeh A., Al Mamlook R.E., Al-Othman A., Tawalbeh M. et al. Breakthrough curves prediction of selenite adsorption on chemically modified zeolite using boosted decision tree algorithms for water treatment applications. *Water.* 2022; 14(16):2519. DOI: 10.3390/w14162519

About the author: **Igor A. Gulshin** — assistant professor; **Moscow State University of Civil Engineering (National Research University) (MGSU)**; 26 Yaroslavskoye shosse, Moscow, 129337, Russian Federation; e-mail: GulshinIA@mgsu.ru.