

УДК 621.314
DOI: 10.18799/24131830/2025/6/5061
Шифр специальности ВАК: 2.4.2
Научная статья

Разработка системы поддержки принятия решений для оценки технического состояния силовых трансформаторов

В.А. Шеломенцев^{1,2}, И.С. Сухачев^{2✉}, С.В. Сидоров²,
В.В. Сушков³, Р.Н. Хамитов², П.В. Чепур²

¹ Тюменский государственный университет, Россия, г. Тюмень

² Тюменский индустриальный университет, Россия, г. Тюмень

³ Нижневартровский государственный университет, Россия, Нижневартовск

✉ suhachevis@tyuiu.ru

Аннотация. Актуальность. Надёжная и экологичная эксплуатация силовых трансформаторов – необходимые требования функционирования современных энергосистем. Важными факторами, приводящими к аварийным ситуациям, являются деградация трансформаторного масла и ненормальные режимы работы электрооборудования. Состав трансформаторного масла служит индикатором технического состояния трансформатора и позволяет оценить ресурс его изоляционных материалов и внутренних компонентов. Своевременная замена масла способствует продлению срока эксплуатации силовых трансформаторов, снижению риска возникновения внезапных отказов и повышению надёжности энергосистемы в целом. В качестве ключевого показателя для выявления ранних признаков износа и потенциальных неисправностей принят прогноз технического состояния силового трансформатора, объединяющий различные параметры, в частности концентрацию растворенных газов и электрические характеристики масла, и позволяющий осуществить прогнозирование его срока эксплуатации. Одним из направлений решения проблем, связанных с определением технического состояния силовых трансформаторов, является применение методов искусственного интеллекта. В связи с этим актуальна разработка систем принятия решений на базе моделей, объединяющих прогнозы классических алгоритмов машинного обучения и моделей, сформированных с использованием методов автоматизированного машинного обучения. Такие системы позволяют совместить преимущества экспертного выбора алгоритмов с возможностями автоматизированного поиска оптимальных структур и параметров модели, что повышает точность оценки технического состояния силового трансформатора и, следовательно, определения на её основе ожидаемого срока его эксплуатации. **Цель:** повышение надёжности силовых трансформаторов при минимизации затрат на техническое обслуживание путём применения методов искусственного интеллекта. **Методы:** статистический анализ хроматографических данных трансформаторного масла; предобработка данных (исключение аномальных и дублирующих записей, z-преобразование); классические методы машинного обучения (линейная регрессия, *Random Forest*, *Extra Trees*, *Hist Gradient Boosting*), валидация модели с разделением выборки данных в соотношении 8:2; разработка структуры модели на базе AutoML с применением специализированной программной платформы FEDOT; расчёт и анализ метрик эффективности модели (R^2 , MAE , MSE , $RMSE$); ансамблирование методами *Averaging*, *Weighted Averaging*, *Stacking*, *Blending* и *XGBoost*. **Результаты.** Разработана ансамблевая модель, предназначенная для комплексной оценки технического состояния силового трансформатора по результатам хроматографического анализа трансформаторного масла и эксплуатационным данным при использовании методов машинного обучения, что позволяет исключить трудоемкие вычисления влияния отдельных признаков и человеческого фактора при выдаче экспертного заключения. Внедрение разработанной модели позволяет объективизировать результаты по оценке остаточного ресурса силового трансформатора и обоснованно перейти к их риск-ориентированному обслуживанию, что, в свою очередь, сокращает эксплуатационные затраты и минимизирует риск возникновения отказа электрооборудования.

Ключевые слова: силовой трансформатор, трансформаторное масло, хроматографический анализ, машинное обучение, регрессионная модель, AutoML, FEDOT, ансамблевая модель, комплексная оценка технического состояния

Благодарности: Статья подготовлена в рамках реализации гранта ФГБОУ ВО «Тюменский индустриальный университет» (проект «Программный комплекс поддержки принятия решений по обеспечению надёжности и устойчивости функционирования электротехнической системы нефтяного промысла»).

Для цитирования: Разработка системы поддержки принятия решений для оценки технического состояния силовых трансформаторов / В.А. Шеломенцев, И.С. Сухачев, С.В. Сидоров, В.В. Сушков, Р.Н. Хамитов, П.В. Чепур // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2025. – Т. 336. – № 6. – С. 120–135. DOI: 10.18799/24131830/2025/6/5061

UDC 621.314
DOI: 10.18799/24131830/2025/6/5061
Scientific paper

Development of a decision support system for assessing the technical condition of power transformers

V.A. Shelomentsev^{1,2}, I.S. Sukhachev²✉, S.V. Sidorov²,
V.V. Sushkov³, R.N. Khamitov², P.V. Chepur²

¹ University of Tyumen, Tyumen, Russian Federation

² Industrial University of Tyumen, Tyumen, Russian Federation

³ Nizhnevartovsk State University, Nizhnevartovsk, Russian Federation

✉suhachevis@tyuiu.ru

Abstract. Relevance. Reliable and environmentally safe operation of power transformers is an essential requirement for the functioning of modern power systems. Transformer oil degradation and abnormal operating conditions of electrical equipment are key factors leading to emergency situations. The composition of transformer oil serves as an indicator of the technical condition of a transformer and enables assessment of the lifespan of its insulating materials and internal components. Timely replacement of oil contributes to extending the operational lifetime of power transformers, reducing the risk of sudden failures, and enhancing overall reliability of the power system. Forecasting the technical condition of a power transformer, integrating various parameters – such as dissolved gas concentrations and electrical characteristics of the oil – is accepted as a crucial indicator for identifying early signs of wear and potential malfunctions, and allows prediction of the transformer operational lifespan. One of the approaches to addressing challenges in determining the technical condition of power transformers involves the application of artificial intelligence methods. In this context, the development of model-based decision-making systems that integrate predictions from classical machine learning algorithms and models generated using automated machine learning techniques is highly relevant. Such systems combine the advantages of expert-driven algorithm selection with the capabilities of automated searches for optimal model architectures and hyperparameters. This hybrid approach enhances the accuracy of assessing a power transformer technical condition and, consequently, improves the determination of its expected service life based on the evaluation. **Aim.** To improve the reliability of power transformers while minimizing maintenance costs through the application of artificial intelligence methods. **Methods.** Statistical analysis of chromatographic data of transformer oil; data preprocessing (elimination of anomalous and duplicate records, z-transformation); classical machine learning methods (linear regression, Random Forest, Extra Trees, Hist Gradient Boosting), model validation using an 8:2 data split; development of a model structure based on AutoML with the specialized FEDOT software platform; calculation and analysis of model performance metrics (R^2 , MAE, MSE, RMSE); ensemble methods *Averaging*, *Weighted Averaging*, *Stacking*, *Blending* and *XGBoost*. **Results.** An ensemble model was developed for the comprehensive assessment of the technical condition of power transformers based on transformer oil chromatography analysis and operational data, using machine learning methods. This approach eliminates labor-intensive calculations of the effect of individual parameters and reduces human factor impact during expert evaluations. Implementation of the proposed model allows objective estimation of the remaining lifespan of power transformers and justifies the transition to risk-oriented maintenance, thereby reducing operational costs and minimizing the risk of electrical equipment failure.

Keywords: power transformer, transformer oil, chromatographic analysis, machine learning, regression model, AutoML, FEDOT, ensemble model, comprehensive assessment of technical condition

Acknowledgements: The article was prepared as part of the implementation of a grant from Industrial University of Tyumen for the project titled «Decision Support Software Complex for Ensuring the Reliability and Stability of Oilfield Electrical System Operation».

For citation: Shelomentsev V.A., Sukhachev I.S., Sidorov S.V., Sushkov V.V., Khamitov R.N., Chepur P.V. Development of a decision support system for assessing the technical condition of power transformers. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2025, vol. 336, no. 6, pp. 120–135. DOI: 10.18799/24131830/2025/6/5061

Введение

Надёжная и экологичная эксплуатация силовых трансформаторов является важнейшим аспектом функционирования современных электроэнергети-

ческих систем. Одной из основных причин аварийности трансформаторов являются ненормальные режимы работы электрооборудования и деградация трансформаторного масла, которое служит индикатором

тором состояния трансформатора, предоставляя информацию о состоянии изоляционных материалов и внутренних компонентов устройства [1–5].

Актуальность исследования обусловлена высокой аварийностью и значительной стоимостью силовых трансформаторов в совокупности с существенным старением трансформаторного парка. Своевременная замена трансформаторного масла напрямую влияет на эксплуатационные характеристики трансформатора, затраты на его техническое обслуживание и экологические показатели, способствует продлению срока эксплуатации и снижению риска возникновения внезапных отказов [6].

Комплексная оценка технического состояния (ТС) трансформатора представляет собой показатель, включающий концентрацию растворенных газов, электрические характеристики масла и ряд иных значимых параметров и позволяющий выявлять ранние признаки износа и потенциальных неисправностей [7–9].

Одним из направлений решения проблем, связанных с определением ТС силовых трансформаторов, является применение методов машинного обучения и искусственного интеллекта [3, 10–15]. Например, в работе [3] проведён обзор методов анализа растворенных газов, связанных с традиционной интерпретацией линейных зависимостей, применение которой снижает точность диагностирования при нестандартных режимах работы трансформаторов. В [13] предложено комбинирование алгоритмов машинного обучения для улучшения интерпретации результатов анализа растворенных газов, однако для модели не применяются автоматизированные методы оптимизации, что требует ручного подбора гиперпараметров.

В связи с этим актуальна разработка систем принятия решений на базе моделей, объединяющих прогнозы классических алгоритмов машинного обучения и моделей, сформированных с использованием методов автоматизированного машинного обучения (AutoML). Такой подход позволяет совместить преимущества экспертного выбора алгоритмов с возможностями автоматизированного поиска оптимальных структур и параметров модели, что повышает точность оценки ТС силового трансформатора и, следовательно, определения на её основе ожидаемого срока его эксплуатации.

Целью настоящей работы является повышение надёжности силовых трансформаторов при минимизации затрат на техническое обслуживание путём применения методов искусственного интеллекта, что требует решения следующих задач:

1) описание, предобработка и статистический анализ набора данных, включающего химические измерения, электрические параметры и расчётные индикаторы состояния оборудования;

2) разработка системы принятия решений на базе модели, объединяющей прогнозы классических алгоритмов машинного обучения и моделей, сформированных с помощью автоматизированного подбора структуры и гиперпараметров (AutoML), что позволяет повысить точность оценки ТС трансформатора.

Описание, предобработка и статистический анализ набора данных

В настоящем исследовании использован набор данных, включающий результаты химических измерений, электрические параметры и расчётные индикаторы, характеризующие техническое состояние силовых трансформаторов. Набор данных собран и систематизирован научными коллективами в рамках работ по анализу причин отказов трансформаторов, в том числе с применением методов машинного обучения [2–8, 14]. В качестве исходного массива данных использованы результаты анализа хроматографических данных проб масла силовых трансформаторов [10–12, 14–18], содержащие 470 записей. Записи содержат различные аспекты состояния трансформатора по 16 признакам, приведенным в табл. 1.

Анализ записей табл. 1 позволяет диагностировать техническое состояние силовых трансформаторов, при этом наиболее значимыми являются признаки, связанные с концентрацией растворенных в масле газов. Высокая концентрация растворенного водорода, метана и ацетилена, превышающая нормативные значения, является признаком длительного перегрева трансформатора, возникновения неоднократных электрических разрядов и разложения изоляционных материалов. Согласно МЭК 60599, допустимые концентрации этих газов в изоляционном масле установлены на уровне, не превышающем 100 ppm для водорода, 300 ppm – для метана, 10 ppm – для ацетилена. Повышенная концентрация дибензилдисульфида, превышающая нормативные значения – 50 ppm, согласно МЭК 60599, связана с содержанием сернистых соединений в масле, что повышает риск развития коррозии металла и ускоряет износ оборудования. Коэффициент мощности трансформатора, поверхностное натяжение и диэлектрическая прочность масла позволяют оценить надёжность работы трансформатора. Содержание воды в масле существенно влияет на его диэлектрические свойства и процессы старения изоляции.

На первом этапе для повышения точности оценки ТС силовых трансформаторов по признакам в табл. 1 необходимо производить предобработку данных. Вначале из массива данных удаляются записи с ошибочными или повторяющимися значениями путём сравнения их уникальных идентификаторов и ключевых признаков.

Таблица 1. Описание признаков набора данных

Table 1. Description of data set attributes

Название признака Feature name	Химическая формула Chemical formula	Описание Description	Единицы измерения Units
Водород/Hydrogen	H ₂	Концентрация водорода в изоляционном масле Hydrogen concentration in insulating oil	млн ⁻¹ ppm
Кислород/Oxygen	O ₂	Концентрация кислорода в изоляционном масле Oxygen concentration in insulating oil	
Азот/Nitrogen	N ₂	Концентрация азота в изоляционном масле Nitrogen concentration in insulating oil	
Метан/Methane	CH ₄	Концентрация метана в изоляционном масле Methane concentration in insulating oil	
Моноксид углерода Carbon monoxide	CO	Концентрация монооксида углерода в изоляционном масле Carbon monoxide concentration in insulating oil	
Диоксид углерода Carbon dioxide	CO ₂	Концентрация диоксида углерода в изоляционном масле Carbon dioxide concentration in insulating oil	
Этилен/Ethylene	C ₂ H ₄	Концентрация этилена в изоляционном масле Ethylene concentration in insulating oil	
Этан/Ethane	C ₂ H ₆	Концентрация этана в изоляционном масле Ethane concentration in insulating oil	
Ацетилен/Acetylene	C ₂ H ₂	Концентрация ацетилена в изоляционном масле Acetylene concentration in insulating oil	
Дибензилдисульфид Dibenzyl disulfide	DBDS	Концентрация дибензилдисульфида в изоляционном масле Dibenzyl disulfide concentration in insulating oil	
Коэффициент мощности трансформатора Power factor	–	Отношение активной мощности к полной мощности Ratio of active power to apparent power	отн. ед. arb. units
Поверхностное натяжение Interfacial tension	–	Поверхностное натяжение на границе раздела фаз трансформаторных масел с водой Interfacial tension at the phase boundary between transformer oils and water	Н/м N/m
Диэлектрическая прочность Dielectric rigidity	–	Способность масла выдерживать электрическое напряжение без пробоя Ability of oil to withstand electrical voltage without breakdown	кВ/мм kV/mm
Содержание воды Water content	–	Концентрация воды в изоляционном масле Water concentration in insulating oil	млн ⁻¹ ppm
Индекс состояния трансформатора Health index	–	Комплексный показатель состояния трансформатора Comprehensive indicator of a transformer health	отн. ед. arb. units
Ожидаемый срок эксплуатации Life expectation	–	Прогнозируемый срок эксплуатации Predicted remaining lifetime	лет years

Затем осуществляется проверка типов данных, при которой все признаки преобразуются в числовой тип с использованием методов библиотеки *pandas* [19], что необходимо для обеспечения корректной обработки данных алгоритмами машинного обучения [13]. Завершающим этапом предобработки является стандартизация данных, заключающаяся в приведении признаков к единому масштабу с нулевым средним и единичным стандартным отклонением посредством z-преобразования:

$$X_{\text{ст}} = \frac{X - \mu}{\sigma},$$

где $X_{\text{ст}}$ – стандартизированное значение признака; X – значение признака из выборки; μ – среднее значение признака по выборке; σ – стандартное отклонение признака.

Выполнение стандартизации обеспечивает совместимость с алгоритмами машинного обучения, которые требуют числового ввода и данных в определенном масштабе для выполнения математических операций.

Вторым этапом выполняется первичный статистический анализ, рассчитываются основные статистические показатели для каждого признака: среднее значение, медиана, стандартное отклонение, минимальные и максимальные значения, с целью получения информации о распределении данных, выявления аномалий и определения характера исследуемых признаков. Дескриптивный анализ каждого признака выполняется независимо друг от друга; в качестве примера результаты анализа по содержанию азота приведены на рис. 1.

Так как модели машинного обучения чувствительны к распределению данных, подготовка признаков с учётом их формы распределения улучшает сходимость моделей и стабильность их результатов работы [20, 21]. Оценка гистограммы помогает выявить небаланс данных, что особенно критично для категориальных задач. Диаграмма размаха позволяет обнаружить и обработать выбросы, способные исказить статистические показатели и ухудшить производительность моделей машинного обучения, особенно регрессионных [22].

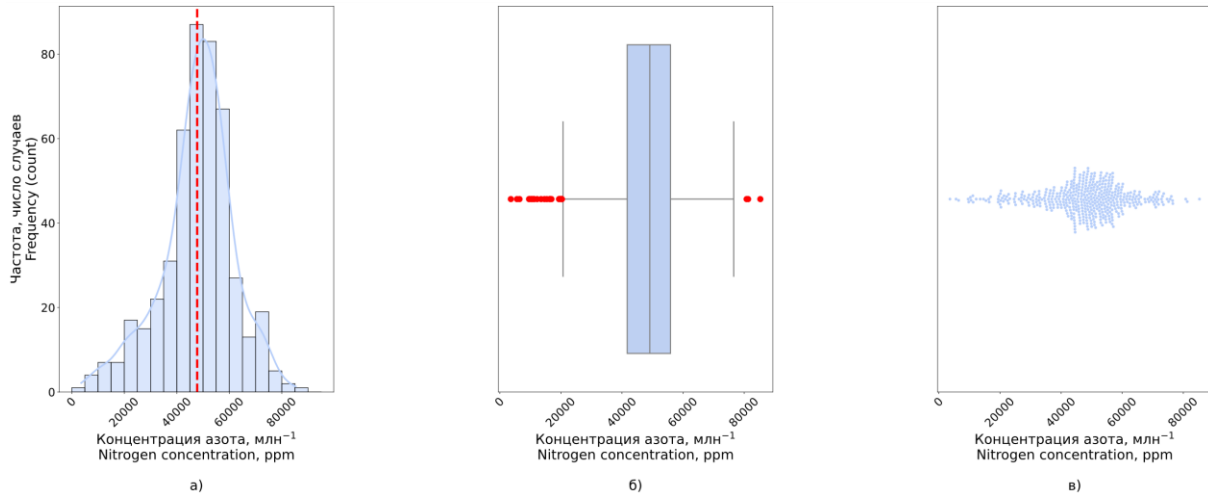


Рис. 1. Графики дескриптивного анализа содержания азота: а) гистограмма распределения; б) диаграмма размаха; в) плотностный график
Fig. 1. Descriptive analysis charts: а) nitrogen distribution histogram; б) nitrogen boxplot; в) nitrogen density plot

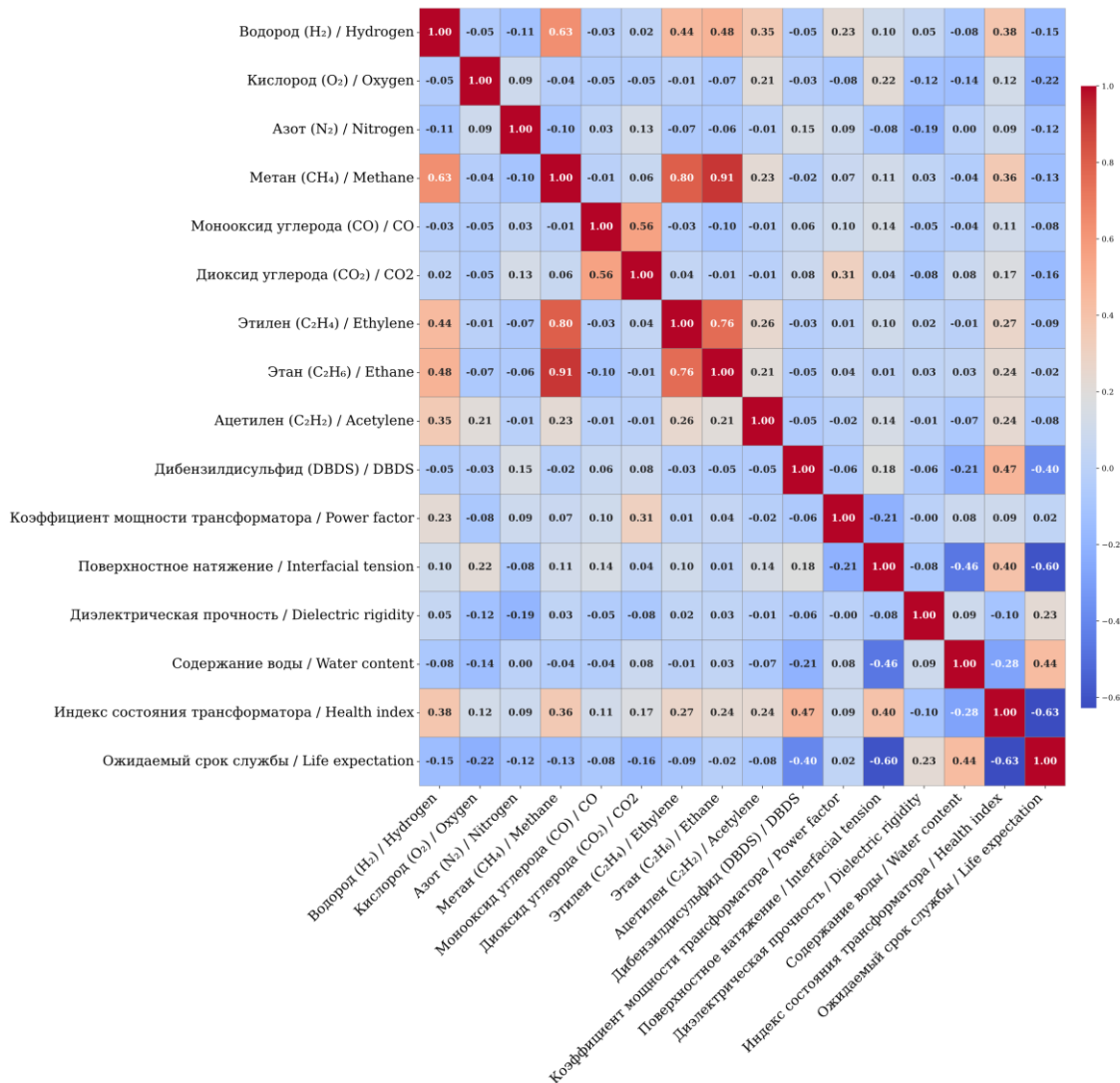


Рис. 2. Корреляционная матрица признаков набора данных для анализа взаимосвязей
Fig. 2. Correlation matrix of dataset features for data analysis

Плотностный график отображает каждое наблюдение как отдельную точку, позволяя обнаружить кластеры и концентрации данных, выполнить поиск паттернов и аномалий, выявить скрытые сегменты в данных, что способствует решению задач кластеризации и сегментации. Таким образом, дескриптивный анализ исходных данных необходим для построения надёжной модели машинного обучения, способной предсказывать ТС силовых трансформаторов с высокой точностью и устойчивостью к аномалиям.

Заключительным этапом с целью определения взаимосвязей между признаками и их влияния на целевую переменную – прогноз ТС силового трансформатора – необходимо выполнить корреляционный анализ данных табл. 1, в результате которого построена корреляционная матрица (рис. 2), отображающая коэффициенты корреляции Пирсона между признаками.

Анализ корреляционной матрицы показал, что существует сильная положительная корреляция целевой переменной с концентрацией следующих газов: водородом, метаном и этиленом, что согласуется с результатами хроматографических исследований.

Таким образом, имеется взаимосвязь концентрации перечисленных растворенных газов и определенных неисправностей трансформатора. Подтверждено существование отрицательной корреляции между ТС силового трансформатора и его ожидаемым сроком эксплуатации, так как ухудшение состояния оборудования сопровождается сокращением его остаточного ресурса.

Оценка целесообразности обработки выбросов при прогнозировании ТС силового трансформатора

Полученные в результате дескриптивного анализа диаграммы размаха содержат значительные выбросы данных в ряде признаков (например, на рис. 1, б показаны красным цветом), поэтому необходимо проанализировать их влияние на качество модели прогнозирования и оценки ТС силового трансформатора и обосновать выбор оптимальных методов обработки данных.

Для оценки качества разработанных моделей прогнозирования технического состояния силового трансформатора (целевой переменной) применяются следующие метрики [23–26]:

1. *Коэффициент детерминации* R^2 отражает способность модели достоверно описывать изменения целевой переменной и позволяет оценить, на сколько модель снижает неопределённость прогнозируемых значений, корректно учитывая данные наблюдений:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

где y_i – истинное значение целевой переменной, $\hat{y}_i$ – прогнозируемое моделью значение, $\bar{y}$ – среднее значение целевой переменной, n – количество наблюдений.

Чем ближе значение коэффициента детерминации R^2 к 1, тем полнее модель учитывает вариативность данных.

2. *Среднеквадратичная ошибка MSE (Mean Squared Error)* отражает точность предсказания модели и измеряет средний квадрат отклонений прогнозируемого значения от истинного значения целевой переменной:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

Возведение отклонений в квадрат увеличивает чувствительность MSE к большим ошибкам (выбросам данных). Чем меньше MSE , тем выше точность модели.

3. *Средняя абсолютная ошибка MAE (Mean Absolute Error)* определяет среднее абсолютное отклонение прогнозируемого моделью значения от истинного значения целевой переменной:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|.$$

MAE измеряется в исходных единицах целевой переменной, что позволяет оценить типичную величину ошибки без дополнительного усиления влияния выбросов данных. Чем ниже MAE , тем меньше величина ошибок прогнозирования модели.

4. *Корень из среднеквадратичной ошибки RMSE (Root Mean Squared Error)* определяется на основе MSE и позволяет получить ошибку в тех же единицах измерения, что и целевая переменная, сохраняя чувствительность к выбросам данных:

$$RMSE = \sqrt{MSE}.$$

Чем меньше значение $RMSE$, тем выше точность модели.

Применение приведённых метрик позволяет оценить как общую способность модели объяснять вариативность исходных данных, так и величину ошибок прогнозирования целевой переменной.

Отметим, что при значениях коэффициента детерминации R^2 ниже 0,5 модель характеризуется низкой достоверностью описания изменения целе-

вой переменной, что свидетельствует о существенных недостатках в построении предсказательной зависимости. В диапазоне значений R^2 от 0,5 до 0,6 модель демонстрирует недостаточную, а при значениях от 0,6 до 0,7 – умеренную объясняющую способность; при этом в обоих случаях требуется ее оптимизация. При значениях R^2 от 0,7 до 0,8 достоверность модели считается приемлемой для предварительного анализа данных. С дальнейшим повышением R^2 модель охватывает большую часть вариативности данных, что свидетельствует о возрастании достоверности её прогнозов [23, 26]. Таким образом, для практического применения модели прогнозирования ТС силового трансформатора требуется обеспечить значение коэффициента детерминации R^2 не ниже 0,8, что означает учёт не менее 80 % вариативности исходных данных. В условиях сопоставимых значений коэффициента детерминации целесообразно отдавать предпочтение моделям с минимальными значениями метрик MSE , MAE и $RMSE$ вследствие меньшего количества абсолютных ошибок предсказаний и, следовательно, более высокой точности оценки ТС силового трансформатора.

В рамках исследования решается задача оценки ТС силового трансформатора на основе комплексного набора диагностических данных, включающего показатели его рабочих режимов, химический состав трансформаторного масла и электрические параметры. Поскольку целевая переменная имеет количественный непрерывный характер, отражающий степень деградации, перегрева и другие ключевые индикаторы состояния трансформатора, для её оценки целесообразно применение методов регрессии, позволяющих установить математическую зависимость с входными признаками, оценить влияние отдельных признаков из табл. 1 и обеспечить интерпретируемость прогноза ТС трансформатора.

В качестве базовой модели для первоначального анализа выбран классический метод линейной регрессии [27–30] вследствие относительной простоты реализации, минимального использования вычислительных ресурсов и достаточной интерпретируемости результатов на первичном этапе обработки диагностических данных. Благодаря своей структуре модель линейной регрессии обеспечивает возможность быстрого установления первоначальной зависимости между диагностическими показателями и ТС силового трансформатора, что является важным этапом в условиях работы с ограниченным объёмом разнородных исходных данных, имеющих различную размерность. Для построения модели весь массив данных разделён на обучающую и тестовую выборки в соотношении 8:2 с фиксированным значением *random_state*, что позволяет проводить обучение на основной части данных и объективно оцени-

вать обобщающую способность модели на независимой выборке, а также обеспечивает воспроизводимость результатов при сравнении различных моделей [31, 32]. Кроме того, использование линейной регрессии позволяет сформировать базовую линию, предназначенную для сравнения результатов более сложных моделей, если линейная зависимость в определенной мере объясняет вариативность целевой переменной. Анализ результатов базовой модели позволяет выявить потенциальные проблемы в исходных данных, что является необходимым этапом перед переходом к более сложным алгоритмам прогнозирования.

В процессе сбора данных хроматографического анализа масла для оценки ТС силовых трансформаторов выбросы могут быть обусловлены как техническими погрешностями измерений, так и эксплуатационными факторами, связанными с различными режимами работы электрооборудования. Неточности могут возникать вследствие калибровочных ошибок измерительных приборов, электромагнитных помех, сбоев в передаче данных или программного обеспечения, обрабатывающего сигналы системы мониторинга. В то же время значительные отклонения параметров могут быть следствием специфических эксплуатационных условий, включая аварийные режимы, нестабильные нагрузки, переходные процессы при коммутациях, перегрев активных частей трансформатора, а также влияние внешних факторов, таких как колебания температуры, влажности или механические воздействия. В отдельных случаях выбросы данных могут быть вызваны редкими аномальными событиями, например грозовыми перенапряжениями или ненормальными режимами работы системы электропитания. Корректная идентификация и анализ подобных выбросов являются критически важными для оценки состояния оборудования, поскольку не все аномалии следует рассматривать как погрешности – некоторые из них могут являться индикаторами потенциальных неисправностей или предаварийных состояний.

Одним из эффективных подходов к идентификации выбросов является поиск границ нормальных значений каждого признака методом межквартильного размаха (Interquartile Range, *IQR*) [20].

Метод межквартильного размаха представляет собой статистический инструмент, основанный на квартильных характеристиках выборки, который позволяет определить центральную тенденцию распределения данных и выявить экстремальные значения без предположения о форме распределения. В отличие от параметрических методов, предполагающих нормальность данных, *IQR* основывается исключительно на порядковых характеристиках (первом Q_1 и третьем Q_3 квартилях), что делает

его универсальным и устойчивым к выбросам. Это свойство метода широко применяется в исследованиях, посвящённых диагностике технического состояния оборудования, где данные могут характеризоваться асимметричным или мультимодальным распределением, а также содержать выбросы, обусловленные перечисленными ранее факторами. Обработка таких данных позволяет учитывать возможные отклонения в данных и обеспечивает корректность последующего анализа диагностических параметров. В контексте диагностики технического состояния силовых трансформаторов применение метода *IQR* позволяет объективно отделить основную массу данных от редких, но потенциально информативных выбросов, что обеспечивает корректное вычисление статистических характеристик и повышает качество последующего моделирования. Согласно методу, вычисляют значения первого Q_1 и третьего Q_3 квартилей, после чего рассчитывается межквартильный размах $IQR=Q_3-Q_1$. Для выявления выбросов устанавливаются две границы: нижняя $H_{LB}=Q_1-1,5 \cdot IQR$ и верхняя $H_{UB}=Q_3+1,5 \cdot IQR$. Все значения, выходящие за эти пределы, классифицируются как выбросы и могут быть либо сохранены для дальнейшего анализа влияния аномальных значений на модель, либо отфильтрованы.

Таблица 2. Пороговые значения признаков для выявления выбросов

Table 2. Threshold values of features for outlier detection

Название признака Feature name	Q_1	Q_3	IQR	H_{LB}	H_{UB}
Водород/Hydrogen	4	34	30	-41	79
Кислород/Oxygen	493	15115	14622	-21440	37048
Азот/Nitrogen	41700	55650	13950	20775	76575
Метан/Methane	2	5	3	-2,5	9,5
Моноксид углерода Carbon monoxide	52	359	307	-408	818
Диоксид углерода Carbon dioxide	609	1980	1371	-1448	4037
Этилен/Ethylene	0	4	4	-6	10
Этан/Ethane	0	78	78	-117	195
Ацетилен/Acetylene	0	0	0	0	0
Дибензилдисульфид Dibenzyl disulfide	0	0	0	0	0
Коэффициент мощности трансформатора Power factor	1	1	0	0	2
Поверхностное натяжение Interfacial tension	32	38	6	23	47
Диэлектрическая прочность Dielectric rigidity	52	57	5	44	64
Содержание воды Water content	11	24	13	-9	44

В табл. 2 указаны пороговые значения признаков, определенные на основе метода *IQR*; форми-

рование наборов данных с учётом выбросов и без них проводилось с использованием функции, реализующей метод *IQR*, на основе библиотеки *pandas* в языке программирования *Python*.

Анализ табл. 2 показал, что признаки набора данных существенно различаются по уровню *IQR*. Вопрос о диагностической ценности признаков «Ацетилен» и «Дибензилдисульфид» с $IQR=0$ требует дополнительного анализа. Высокие значения *IQR* (более 500) признаков «Кислород», «Азот» и «Диоксид углерода» означает широкий диапазон наблюдаемых значений, что может быть обусловлено различиями в эксплуатационных режимах, внешними факторами или присутствием выбросов. Небольшие изменения факторов могут приводить к существенным колебаниям данных, следовательно, их корректная интерпретация требует дополнительных исследований и учёта сопутствующих признаков.

Таким образом, применение метода *IQR* позволило установить границы нормальных значений для каждого признака и выделить потенциальные выбросы. С целью влияния выбросов данных на точность оценки ТС силового трансформатора и целесообразность фильтрации исходных данных для обучения модели линейной регрессии на основе табл. 1, 2 подготовлены два набора данных:

- 1) *данные с выбросами* – исходный набор данных без изменений;
- 2) *данные без выбросов* – набор данных после фильтрации выбросов методом *IQR*.

Результаты моделирования с использованием обоих наборов данных приведены в табл. 3.

Таблица 3. Оценка влияния выбросов на качество прогнозирования модели линейной регрессии

Table 3. Assessment of the impact of outliers on the prediction quality of the linear regression model

Наименование метрики Metric name	С выбросами With outliers	Без выбросов Without outliers
R^2	0,597	0,306
MSE	126,362	28,255
MAE	8,116	2,973
$RMSE$	11,241	5,315

Анализ табл. 3 показал, что базовая модель линейной регрессии обладает недостаточной способностью описывать сложные нелинейные взаимосвязи диагностических данных и целевой переменной. При обучении модели на наборе данных с выбросами коэффициент детерминации $R^2=0,597$, т. е. означает учёт менее 60 % вариативности исходных данных. Применение фильтрации выбросов привело к существенному снижению метрик ошибок MSE , MAE и $RMSE$, однако также сопровождалось уменьшением

коэффициента детерминации до $R^2=0,306$, что свидетельствует о потере значимой информации, отражающей в том числе критические (предаварийные) состояния трансформатора. Таким образом, с целью повышения точности оценки ТС силового трансформатора целесообразно не проводить фильтрацию исходных данных для обучения модели регрессии.

Разработка модели оценки ТС трансформатора на базе классических методов машинного обучения

Исходя из требования обеспечить значение коэффициента детерминации R^2 не ниже 0,8 для практического применения модели оценки ТС силового трансформатора следует, что необходимо перейти от базовой модели линейной регрессии к моделям на базе классических методов машинного обучения, способных учитывать нелинейный характер диагностических признаков для повышения точности и устойчивости предсказаний. В результате проведенного сравнительного анализа классических методов машинного обучения с использованием кросс-валидации [33, 34], обеспечивающей объективную оценку обобщающей способности моделей, получена выборка из трех наиболее перспективных моделей: *Extra Trees Regressor*, *Hist Gradient Boosting Regressor* и *Random Forest Regressor*, метрики которых сведены в табл. 4.

Таблица 4. Метрики оценки качества регрессионных моделей на базе классических методов машинного обучения

Table 4. Quality metrics of regression models based on classical machine learning methods

Наименование регрессионных моделей Regression models nomenclature	R^2	MAE	MSE	RMSE
Extra Trees Regressor	0,76	5,21	74	8,59
Hist Gradient Boosting Regressor	0,73	5,50	80	8,94
Random Forest Regressor	0,72	5,80	85	9,22

Анализ табл. 4 показал, что применение перечисленных моделей позволяет добиться более высоких значений коэффициента детерминации R^2 по сравнению с базовой моделью линейной регрессии при приемлемом уровне ошибок по метрикам MAE, MSE и RMSE. Тем не менее величина R^2 не достигла целевого значения, что объясняется рядом причин, среди которых неполнота или шумность исходных данных, недостаточная информативность отдельных признаков, ограниченные возможности классических моделей при моделировании нелинейных процессов.

Таким образом, для повышения точности оценки ТС силового трансформатора целесообразно перейти к моделям, способным более полно учиты-

вать нелинейные взаимосвязи между признаками. Тем не менее известно [35, 36], что недостатками таких моделей на базе алгоритмов машинного обучения является использование большого перечня методов, обладающих значительной вычислительной сложностью и требующих настройки множества гиперпараметров, что в совокупности приводит к существенному увеличению числа возможных структур. Следовательно, целесообразно применение автоматизированного метода отбора и комбинирования моделей, позволяющего ускорить экспериментальные циклы разработки структуры модели и повысить воспроизводимость результатов.

Разработка модели оценки ТС силового трансформатора автоматизированным методом AutoML

Формирование структуры и подбор гиперпараметров модели оценки ТС силового трансформатора реализовано автоматизированным методом (AutoML) с использованием программной платформы FEDOT, опирающейся на эволюционный алгоритм для объединения и настройки этапов предобработки данных и алгоритмов машинного обучения [35, 36]. В отличие от традиционных подходов к поиску гиперпараметров модели [12, 13] эволюционный алгоритм осуществляет последовательное улучшение уже отфильтрованных структур, оптимизируя модели с наиболее высокими коэффициентами детерминации R^2 .

Разработка модели оценки ТС трансформатора выполнена с применением пресета *best_quality*, ориентированного на максимизацию качества модели при решении задачи машинного обучения путём увеличения числа итераций оптимизации и применения расширенных наборов доступных алгоритмов и их гиперпараметров. На рис. 3 приведена оптимизированная структура модели оценки ТС силового трансформатора, сформированная с применением программной платформой FEDOT и состоящая из пяти этапов обработки данных и построения модели.

На первом этапе построения модели выполняется балансировка данных, направленная на устранение диспропорции между значениями целевой переменной, что позволяет предотвратить смещение результатов прогнозирования модели в сторону наиболее часто встречающихся значений. На втором этапе осуществляется предварительное моделирование с использованием *Random Forest Regressor*, позволяющее выявить ключевые зависимости между входными признаками и целевой переменной. В результате формируются промежуточная оценка ТС силового трансформатора и информация о значимости признаков, необходимая для последующей обработки данных.

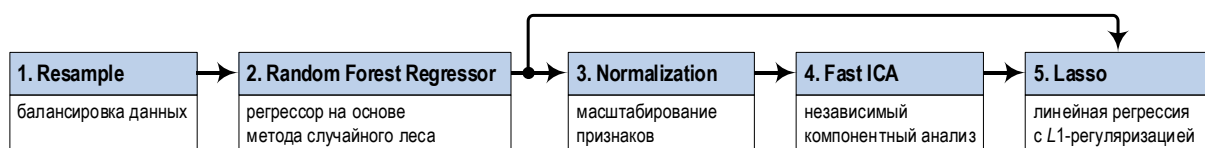


Рис. 3. Оптимизированная структура модели оценки ТС трансформатора, сформированная с применением программной платформы FEDOT

Fig. 3. Optimized structure of the transformer technical condition assessment model formed using the FEDOT software platform

На третьем этапе производится масштабирование признаков, обеспечивающее приведение данных к единому масштабу, что необходимо для корректной работы алгоритмов, чувствительных к диапазонам значений. Четвертым этапом выполняется независимый компонентный анализ, позволяющий снизить размерность данных за счет удаления избыточной информации и устранения коррелированных признаков, что способствует повышению достоверности модели. Заключительным, пятым, этапом моделью линейной регрессии с $L1$ -регуляризацией выполняется окончательная коррекция оценки ТС силового трансформатора, включая незначимые параметры.

Особенностью структуры, приведённой на рис. 3, является параллельное использование выходных данных модели *Random Forest Regressor*, что позволяет улучшить итоговое качество прогнозирования. Разделение потоков данных способствует повышению интерпретируемости модели и сокращению вычислительных затрат за счёт раннего исключения нерелевантных факторов.

Метрики AutoML модели прогнозирования с оптимизированной структурой на базе программной платформы FEDOT приведены в табл. 5.

Таблица 5. Метрики модели AutoML на базе программной платформы FEDOT

Table 5. Quality metrics of the AutoML model based on the FEDOT platform

Наименование регрессионной модели Regression models nomenclature	R^2	MAE	MSE	RMSE
AutoML на базе программной платформы FEDOT AutoML model based on the FEDOT platform	0,771	6,019	79,910	8,937

Таким образом, оптимизированная структура модели, сформированная с использованием программной платформы FEDOT, характеризуется значением коэффициента детерминации $R^2=0,771$, что остается ниже целевого уровня $R^2 \geq 0,8$, необходимого для практического применения полученной оценки ТС силового трансформатора. Результаты

моделирования указывают на ограничения как классических, так и автоматизированных методов в интерпретации сложных взаимосвязей диагностических данных. Для дальнейшего повышения точности прогнозирования целесообразно разработать ансамблевую модель, основанную на интеграции двух независимых подходов, включающих классические методы машинного обучения и автоматизированные методы оптимизации структуры модели, что минимизирует их индивидуальные недостатки и компенсирует ошибки за счет взаимного дополнения прогнозов [13].

Разработка ансамблевой модели комплексной оценки ТС силового трансформатора

Разработка ансамблевой модели комплексной оценки ТС силового трансформатора выполнена в два этапа.

На первом этапе проанализированы результаты классических алгоритмов машинного обучения и выбрана эффективная для дальнейшего применения модель *Extra Trees Regressor*, обладающая наибольшим значением коэффициента детерминации R^2 согласно табл. 4. Далее для повышения точности предсказаний и обобщающей способности выполнена оптимизация модели, заключающаяся в настройке её гиперпараметров с использованием метода перебора по сетке в сочетании с кросс-валидацией [33, 35–37]. Параллельно на первом этапе автоматизированным методом (AutoML) с использованием программной платформы FEDOT сформирована структура модели и подобраны её гиперпараметры.

Вторым этапом для компенсации недостатков базовых моделей за счёт их взаимного дополнения и повышения достоверности прогнозов разработаны ансамблевые модели с применением методов: *Averaging* (простое усреднение), *Weighted Averaging* (взвешенное усреднение), *Stacking* (сложение моделей), *Blending* (смешивание моделей) и *XGBoost* (градиентный бустинг) [13, 35]. Метрики полученных ансамблевых моделей сведены в табл. 6.

Анализ табл. 6 показал, что разработанная ансамблевая модель на основе метода *Weighted Aver-*

aging характеризуется значением коэффициента детерминации $R^2=0,832$, что соответствует критерию $R^2 \geq 0,8$ и свидетельствует о достаточной способности модели описывать нелинейные зависимости и учитывать специфику условий эксплуатации силовых трансформаторов. Анализ ошибок по метрикам MAE, MSE и RMSE подтвердил, что применение метода *Weighted Averaging* обеспечивает наименьшие отклонения прогнозируемых значений от фактических по сравнению с другими рассмотренными способами ансамблирования.

Таблица 6. Метрики ансамблевых моделей

Table 6. Quality metrics of ensemble models

Наименование метода ансамблирования Ensemble method name	R^2	MAE	MSE	RMSE
Averaging	0,814	5,294	63,532	7,970
Weighted Averaging	0,832	5,091	60,898	7,803
Stacking	0,799	4,871	68,809	8,295
Blending	0,720	6,646	95,748	9,785
XGBoost	0,795	5,070	70,163	8,376

Полученные с помощью разработанной ансамблевой модели значения комплексной оценки ТС силовых трансформаторов, разбитых по 5 группам, приведены на рис. 4.

Согласно рис. 4, в группе 1 относительная ошибка прогнозирования ТС трансформаторов мала и не превышает 1 %, результаты модели обладают высокой точностью. В группах 2, 3 относительная ошибка прогнозирования ТС не превышает

5 %, точность результатов модели остаётся высокой, возрастает влияние режимных параметров трансформатора и ограниченность обучающего набора данных. В группе 4 относительная ошибка прогнозирования ТС достигает 9 %, что связано со значительным числом выбросов данных, которые, в частности, могут привести к недостоверной интерпретации критериев табл. 1, точность результатов модели остаётся удовлетворительной. Наибольшая ошибка прогнозирования ТС наблюдается в группе 5, включающей трансформаторы в предаварийном состоянии; точность результатов модели остаётся удовлетворительной, коэффициент детерминации $R^2=0,832$ не ниже допустимого значения 0,8.

В зависимости от величины полученной комплексной оценки ТС силового трансформатора предлагаются следующие рекомендации по его техническому обслуживанию, представленные в табл. 7.

Разработанная система принятия решений на базе ансамблевой модели оказывается более «устойчивой» к шумам и выбросам, чем любая из составляющих моделей ансамбля по отдельности, обеспечивает необходимые требованиям надёжности и точности прогнозирования, что позволяет рекомендовать её для применения в комплексах мониторинга и диагностирования технического состояния силовых трансформаторов. Структура разработанной ансамблевой модели приведена на рис. 5.

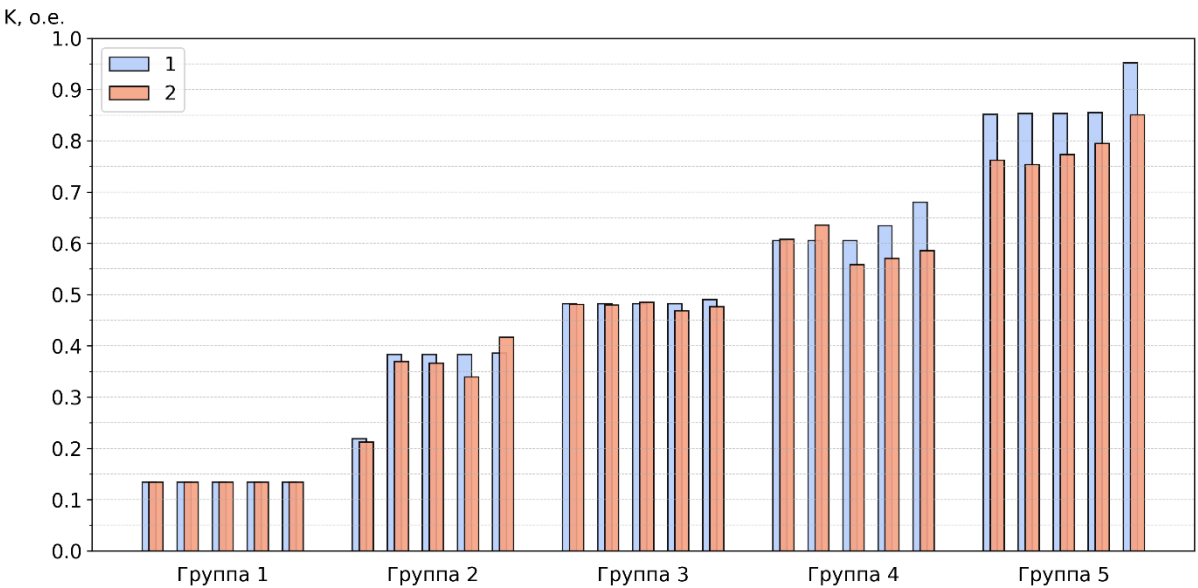


Рис. 4. График комплексной оценки K технического состояния силового трансформатора, где 1 – экспертное значение K, 2 – прогнозируемое значение K

Fig. 4. Graph of the comprehensive assessment K of the technical condition of the power transformer, where 1 – expert value of K, 2 – predicted value of K

Таблица 7. Рекомендации по техническому обслуживанию силового трансформатора

Table 7. Recommendations for power transformer maintenance

Значения комплексной оценки ТС трансформатора Values of the comprehensive assessment of the transformer technical condition	ТС трансформатора Transformer condition	Рекомендации Recommendations
0–0,2	Хорошее Good	Дополнительные мероприятия не требуются No additional measures required
0,2–0,4	Среднее Medium	Периодическое диагностирование ТС Periodic diagnostics of the transformer condition
0,4–0,6	Плохое Bad	Целесообразно применение непрерывного мониторинга параметров трансформатора и проведение текущего ремонта по его результатам Continuous monitoring of transformer parameters is advisable, followed by corrective maintenance based on the results
0,6–0,8	Очень плохое Very bad	Вывод трансформатора в ремонт Transformer decommissioning for repair
0,8–1,0	Критическое Critical	Определение остаточного ресурса трансформатора с последующей его заменой Determination of the remaining lifespan of the transformer followed by replacement

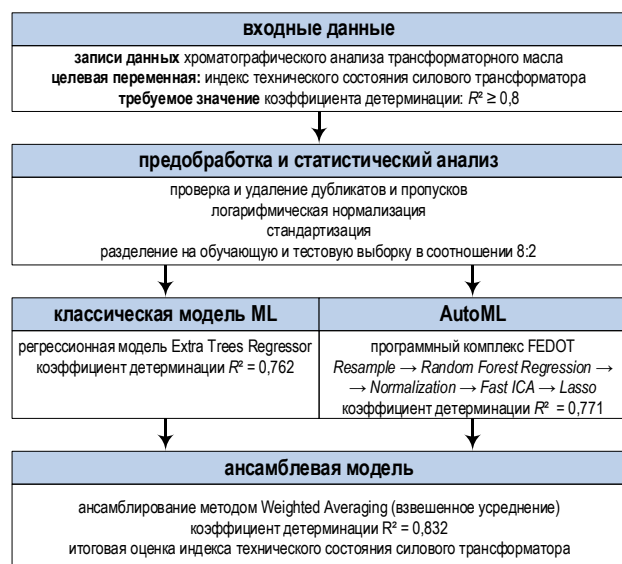


Рис. 5. Структура ансамблевой модели комплексной оценки ТС трансформатора

Fig. 5. Structure of the ensemble model for the comprehensive assessment of the transformer technical condition

Дальнейшее повышение качества прогнозирования возможно за счет расширенной балансировки исходных данных, повышения разнообразия используемых алгоритмов и более детальной настройки гиперпараметров. Интеграция дополнительных искусственных нейронных сетей или усовершенствованных архитектур глубокого машинного обучения в перспективе также позволит повысить способность модели выявлять редкие, но критически важные состояния силовых трансформаторов, что требует дальнейших исследований.

Заключение

1. Обоснована необходимость учёта выбросов при анализе хроматографических данных трансформаторного масла, поскольку экстремальные значения могут не только искажать параметры обучаемых моделей, но и служить индикаторами потенциальных аномалий в работе силового трансформатора.

Использование метода межквартильного размаха позволило подтвердить пороговые значения диагностических параметров, установить границы нормальных и аномальных значений и провести сравнительный анализ влияния выбросов на качество прогнозных моделей.

2. Модели оценки ТС силовых трансформаторов, разработанные на основе классических методов машинного обучения, автоматизированных методов (AutoML) с применением специализированной программной платформы FEDOT требуют дальнейшего усовершенствования вследствие сложной нелинейной структуры входных данных, наличия коррелированных признаков и вариативности условий эксплуатации электрооборудования.
3. Разработанная система поддержки принятия решений на базе ансамблевой модели обеспечивает коэффициент детерминации $R^2=0,832$, превышающий целевое значение, что свидетельствует о достаточной достоверности получаемой комплексной оценки ТС силового трансформатора и допустимости практического применения. Внедрение разработанной системы позволяет использовать методы искусственного интеллекта в системах мониторинга и диагностирования силовых трансформаторов, что позволяет объективизировать результаты по оценке их остаточного ресурса и обоснованно перейти к риск-ориентированному обслуживанию, что в свою очередь сокращает эксплуатационные затраты и минимизирует риск возникновения отказа электрооборудования.

СПИСОК ЛИТЕРАТУРЫ

1. Никифоров И.К. Обзор различных факторов, влияющих на эксплуатационные параметры трансформаторных масел. Контроль параметров трансформаторных масел // Приборы и системы. Управление, контроль, диагностика. – 2021. – № 4. – С. 7–18. DOI: 10.25791/pribor.4.2021.1251.
2. Влияние старения трансформаторного масла на остаточный ресурс трансформатора и его диагностические параметры / А.С. Серебряков, Д.А. Семенов, Е.А. Сбитнев, А.В. Сидорова // Электротехника. – 2023. – № 12. – С. 71–78. DOI: 10.53891/00135860_2023_12_71.
3. Bakar N.A., Abu-Siada A., Islam S. A review of dissolved gas analysis measurement and interpretation techniques // IEEE Electrical Insulation Magazine. – 2014. – Vol. 30. – № 3. – P. 39–49. DOI: 10.1109/MEI.2014.6804740.
4. Яхья А.А., Левин В.М. Оценка эффективности мониторинга технического состояния трансформаторов на основе анализа растворенных газов // Журнал Сибирского федерального университета. Серия: Техника и технологии. – 2020. – Т. 13. – № 4. – С. 438–448. DOI: 10.17516/1999-494X-0235.
5. МЭК 60599:2022. Электрооборудование маслonaполненное в условиях эксплуатации. Руководство по интерпретации результатов анализа растворенных и свободных газов. – Женева: Международная электротехническая комиссия, 2022. – 80 с.
6. Оценка остаточного ресурса изоляции погружного электродвигателя установок электрических центробежных насосов добычи нефти при воздействиях импульсных перенапряжений / В.В. Сушков, В.В. Тимошкин, И.С. Сухачев, С.В. Сидоров // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2017. – Т. 328. – № 10. – С. 74–80.
7. Вахнина В.В., Марков Е.В. Разработка программы для повышения точности диагностики развивающихся дефектов в силовых трансформаторах на основе хроматографического анализа растворенных газов в масле // Известия высших учебных заведений. Электромеханика. – 2023. – Т. 66. – № 3. – С. 36–42. DOI: 10.17213/0136-3360-2023-3-36-42.
8. Идентификация состояния силовых трансформаторов 110(35)/10 кВ на основе анализа данных тепловизионного обследования и методов искусственного интеллекта / И.С. Сухачев, В.А. Шеломенцев, Г.А. Хмара, С.В. Сидоров, Е.И. Попов, А.А. Швецова // Электротехнические и информационные комплексы и системы. – 2024. – Т. 20. – № 2. – С. 22–36. DOI: 10.17122/1999-5458-2024-20-2-22-36.
9. Разработка модели прогнозирования технического состояния нефтепромыслового энергомеханического оборудования с применением искусственной нейронной сети / И.С. Сухачев, С.В. Сидоров, В.В. Сушков, В.Р. Антропова // Электротехнические и информационные комплексы и системы. – 2022. – Т. 18. – № 1. – С. 88–94. DOI: 10.17122/1999-5458-2022-18-1-88-94.
10. Naresh R., Sharma V., Vashisth M. An integrated neural fuzzy approach for fault diagnosis of transformers // IEEE Transactions on Power Delivery. – 2008. – Vol. 23. – № 4. – P. 2017–2024. DOI: 10.1109/TPWRD.2008.2002652.
11. Neuro fuzzy schemes for fault detection in power transformer / V. Duraisamy, N. Devarajan, D. Somasundareswari, A.A.M. Vasanth, S.N. Sivanandam // Applied Soft Computing. – 2007. – Vol. 7. – № 2. – P. 534–539. DOI: 10.1016/j.asoc.2006.10.001.
12. Castro A.R.G., Miranda V. Knowledge discovery in neural networks with application to transformer failure diagnosis // IEEE Transactions on Power Systems. – 2005. – Vol. 20. – № 2. – P. 717–724. DOI: 10.1109/TPWRS.2005.846074.
13. Senoussaoui M.E.A., Brahmi M., Fofana I. Combining and comparing various machine-learning algorithms to improve dissolved gas analysis interpretation // IET Generation, Transmission & Distribution. – 2018. – Vol. 12. – № 16. – P. 3673–3679. DOI: 10.1049/iet-gtd.2018.0059.
14. Arias Velásquez R.M., Mejía Lara J.V. Corrosive sulphur effect in power and distribution transformers failures and treatments // Engineering Failure Analysis. – 2018. – Vol. 92. – P. 240–267. DOI: 10.1016/j.engfailanal.2018.05.018.
15. Arias Velásquez R.M., Mejía Lara J.V., Melgar A. Converting data into knowledge for preventing failures in power transformers // Engineering Failure Analysis. – 2019. – Vol. 101. – P. 215–229. DOI: 10.1016/j.engfailanal.2019.03.027.
16. A probabilistic classifier for transformer dissolved gas analysis with a particle swarm optimizer / W.H. Tang, J.Y. Goulermas, Q.H. Wu, Z.J. Richardson, J. Fitch // IEEE Transactions on Power Delivery. – 2008. – Vol. 23. – № 2. – P. 751–759. DOI: 10.1109/TPWRD.2008.915812.
17. Yadaiah N., Ravi N. Internal fault detection techniques for power transformers // Applied Soft Computing. – 2011. – Vol. 11. – P. 5259–5269. DOI: 10.1016/j.asoc.2011.05.034.
18. Duval M., De Pabla A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases // IEEE Electrical Insulation Magazine. – 2001. – Vol. 17. – № 2. – P. 31–41. DOI: 10.1109/57.917529.
19. Gopal S., Patro K., Sahu K.K. Normalization: a preprocessing stage // International Advanced Research Journal in Science, Engineering and Technology. – 2015. – Vol. 2. – № 3. – P. 20–22. DOI: 10.17148/IARJSET.2015.2305.
20. Mueller A., Schmidt B., Keller C. Evaluating data preprocessing methods and their impact on machine learning performance // Computational Intelligence (De Gruyter). – 2022. – Vol. 20. – № 3. – P. 145–162. DOI: 10.1515/comp-2022-0278.
21. A review of modern data preprocessing techniques for machine learning / L. Wang, X. Li, Y. Chen, Z. Zhang // 2022 IEEE International Conference on Big Data. – 2022. – P. 120–125. DOI: 10.1109/BigData.2022.10714767.
22. Alsharif M., Abujarrad R., Al-Ramahi S. Data preprocessing techniques in Python using Pandas library // 2020 IEEE International Conference on Data Science and Advanced Analytics (DSAA). – 2020. – P. 1–6. DOI: 10.1109/DSAA.2020.10575760.
23. Chicco D., Warrens M.J., Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation // PeerJ Computer Science. – 2021. – Vol. 7. – Article e623. DOI: 10.7717/peerj-cs.623.
24. Comparative analysis of regression metrics for machine learning models / N. Srivastava, M. Agarwal, V. Bhardwaj, T. Lamba // Proceedings of the 2022 International Conference on Advances in Computing, Communication, and Intelligent Systems (ICACITE), 2022. – P. 475–480. DOI: 10.1109/ICACITE53722.2022.9823692.

25. Dewi A.L., Akrom M. Evaluating machine learning model performance using R^2 , RMSE, and MAE // *Jurnal Informatika: Jurnal Pengembangan IT*. – 2024. – Vol. 1. – № 2. – P. 45–53. DOI: 10.62411/jimat.v1i2.11053.
26. Ковригин Б.Л., Лавринович А.В. Применение коэффициента детерминации для сравнения осциллограмм отклика силового трансформатора, полученных при диагностике методом наносекундных низковольтных импульсов // *Известия Томского политехнического университета. Инжиниринг георесурсов*. – 2018. – Т. 329. – № 4. – С. 154–161.
27. Maulud D.H., Abdulazeez A. A comparative review of linear regression and polynomial regression // *Journal of Applied Statistics and Theory*. – 2020. – Vol. 15. – № 2. – P. 45–53. DOI: 10.38094/jastt1457.
28. Linear regression as a baseline model for machine learning / K. Bartol, M. Novak, L. Horvat, I. Zidar // *Sensors*. – 2022. – Vol. 22. – № 5. – P. 188–195. DOI: 10.3390/s22051885.
29. An overview of linear regression in machine learning / A. Radhakrishnan, P. Gupta, S. Kumar, R. Singh // *IEEE Access*. – 2022. – Vol. 10. – P. 15000–15008. DOI: 10.1109/ACCESS.2022.3141592.
30. Improving linear regression interpretability with adaptive methods / L. Munkhdalai, T. Enkhbayar, B. Purev, D. Davaadorj // *Mathematics*. – 2020. – Vol. 8. – № 11. – P. 150–160. DOI: 10.1007/s11293-021-09711-8.
31. Bichri H., Chergui A., Hain M. Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets // *International Journal of Advanced Computer Science and Applications (IJACSA)*. – 2024. – Vol. 15. – № 2. – P. 331–337. DOI: 10.14569/IJACSA.2024.0150235.
32. Xu Z., De A. Assessing model accuracy using random data split: a simulation study // *Journal of Biopharmaceutical Statistics*. – 2022. – Vol. 33. – P. 131–139. DOI: 10.1080/10543406.2022.2089158.
33. Sáez J.A., Romero-Béjar J.L. Impact of k-fold cross-validation on regression model performance under dataset shift // *Mathematics*. – 2022. – Vol. 10. – № 14. – Article 2538. DOI: 10.3390/math10142538.
34. Bootstrapping the cross-validation estimate / B. Cai, F. Pellegrini, M. Pang, C. de Moor, C. Shen, V. Charu, L. Tian // *ISPRS Archives*. – 2022. – Vol. XLVIII-4/W3-2022. – P. 111–118. DOI: 10.5194/isprs-archives-XLVIII-4-W3-2022-111-2022.
35. Automated evolutionary approach for the design of composite machine learning pipelines / N.O. Nikitin, P. Vychuzhanin, M. Sarafanov, I.S. Polonskaia, I. Revin, I.V. Barabanova, G. Maximov, A.V. Kalyuzhnaya, A. Boukhanovsky // *Future Generation Computer Systems*. – 2022. – Vol. 127. – P. 109–125. DOI: 10.1016/j.future.2021.08.022.
36. AutoML for feature selection and model tuning applied to fault severity diagnosis in spur gearboxes / M. Cerrada, L. Trujillo, D. Hernández, H.A. Correa Zevallos, J.C. Macancela, D. Román Cabrera, R.V. Sánchez // *Mathematical and Computational Applications*. – 2022. – Vol. 27. – № 1. – Article 6. DOI: 10.3390/mca27010006.
37. A rank analysis and ensemble machine learning model for load forecasting in the nodes of the central mongolian power system / T. Osgonbaatar, P. Matrenin, M. Safaraliev, I. Zicmane, A. Rusina, S. Kokin // *Inventions*. – 2023. – Vol. 8. – № 5. – P. 114. DOI: 10.3390/inventions8050114.

Информация об авторах

Владислав Александрович Шеломенцев, ассистент Тюменского государственного университета, Россия, 625003, г. Тюмень, ул. Володарского, 6; аспирант кафедры электроэнергетики, Тюменский индустриальный университет, Россия, 625000, г. Тюмень, ул. Володарского, 38. shelomentseva72@mail.ru

Илья Сергеевич Сухачев, кандидат технических наук, доцент, заведующий базовой кафедрой АО «СУЭНКО», Тюменский индустриальный университет, Россия, 625000, г. Тюмень, ул. Володарского, 38. suhachevis@tyuiu.ru

Сергей Владимирович Сидоров, кандидат технических наук, доцент кафедры электроэнергетики, Тюменский индустриальный университет, Россия, 625000, г. Тюмень, ул. Володарского, 38. si-dorovsv@tyuiu.ru

Валерий Валентинович Сушков, доктор технических наук, профессор кафедры энергетики, Нижневартровский государственный университет, Россия, 628605, Нижневартовск, ул. Дзержинского, 11; sushkovvv50@mail.ru

Рустам Нуриманович Хамитов, доктор технических наук, профессор кафедры электроэнергетики Института промышленных технологий и инжиниринга, Тюменский индустриальный университет, Россия, 625000, г. Тюмень, ул. Володарского, 38; apple_27@mail.ru

Петр Владимирович Чепур, кандидат технических наук, доцент, заведующий кафедрой строительной механики, Тюменский индустриальный университет, Россия, 625000, г. Тюмень, ул. Володарского, 38. chepur_p_v@mail.ru

Поступила в редакцию: 19.03.2025

Поступила после рецензирования: 07.04.2025

Принята к публикации: 12.05.2025

REFERENCES

1. Nikiforov I.K. Review of various factors affecting the operational parameters of transformer oils. Control of transformer oil parameters. *Devices and Systems. Control, Monitoring, Diagnostics*, 2021, no. 4, pp. 7–18. (In Russ.) DOI: 10.25791/pribor.4.2021.1251.

2. Serebryakov A.S., Semenov D.A., Sbitnev E.A., Sidorova A.V. Influence of transformer oil aging on the residual life of the transformer and its diagnostic parameters. *Electrical Engineering*, 2023, no. 12, pp. 71–78. (In Russ.) DOI: 10.53891/00135860_2023_12_71.
3. Bakar N.A., Abu-Siada A., Islam S. A review of dissolved gas analysis measurement and interpretation techniques. *IEEE Electrical Insulation Magazine*, 2014, vol. 30, no. 3, pp. 39–49. DOI: 10.1109/MEI.2014.6804740.
4. Yakhya A.A., Levin V.M. Evaluation of the efficiency of transformer condition monitoring based on dissolved gas analysis. *Journal of Siberian Federal University. Engineering & Technologies*, 2020, vol. 13, no. 4, pp. 438–448. (In Russ.) DOI: 10.17516/1999-494X-0235.
5. IEC 60599:2022. *Mineral oil-filled electrical equipment in service – Guidance on the interpretation of dissolved and free gases analysis*. Geneva, International Electrotechnical Commission, 2022. 80 p.
6. Sushkov V.V., Timoshkin V.V., Sukhachev I.S., Sidorov S.V. Assessment of the residual insulation life of a submersible electric motor for electrical centrifugal pump installations under impulse overvoltages. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2017, vol. 328, no. 10, pp. 74–80. (In Russ.)
7. Vakhnina V.V., Markov E.V. Development of a program to improve the accuracy of diagnosing developing defects in power transformers based on chromatographic analysis of dissolved gases in oil. *Proceedings of Higher Educational Institutions. Electromechanics*, 2023, vol. 66, no. 3, pp. 36–42. (In Russ.) DOI: 10.17213/0136-3360-2023-3-36-42.
8. Sukhachev I.S., Shelomentsev V.A., Khmara G.A., Sidorov S.V., Popov E.I., Shvetsova A.A. Identification of the condition of 110(35)/10 kV power transformers based on thermal imaging survey data analysis and artificial intelligence methods. *Electrical and Information Complexes and Systems*, 2024, vol. 20, no. 2, pp. 22–36. (In Russ.) DOI: 10.17122/1999-5458-2024-20-2-22-36.
9. Sukhachev I.S., Sidorov S.V., Sushkov V.V., Antropova V.R. Development of a model for predicting the technical condition of oilfield electromechanical equipment using artificial neural networks. *Electrical and Information Complexes and Systems*, 2022, vol. 18, no. 1, pp. 88–94. (In Russ.) DOI: 10.17122/1999-5458-2022-18-1-88-94.
10. Naresh R., Sharma V., Vashisth M. An integrated neural fuzzy approach for fault diagnosis of transformers. *IEEE Transactions on Power Delivery*, 2008, vol. 23, no. 4, pp. 2017–2024. DOI: 10.1109/TPWRD.2008.2002652.
11. Duraisamy V., Devarajan N., Somasundareswari D., Vasanth A.A.M., Sivanandam S.N. Neuro fuzzy schemes for fault detection in power transformer. *Applied Soft Computing*, 2007, vol. 7, no. 2, pp. 534–539. DOI: 10.1016/j.asoc.2006.10.001.
12. Castro A.R.G., Miranda V. Knowledge discovery in neural networks with application to transformer failure diagnosis. *IEEE Transactions on Power Systems*, 2005, vol. 20, no. 2, pp. 717–724. DOI: 10.1109/TPWRS.2005.846074.
13. Senoussaoui M.E.A., Brahmi M., Fofana I. Combining and comparing various machine-learning algorithms to improve dissolved gas analysis interpretation. *IET Generation, Transmission & Distribution*, 2018, vol. 12, no. 16, pp. 3673–3679. DOI: 10.1049/iet-gtd.2018.0059.
14. Arias Velásquez R.M., Mejía Lara J.V. Corrosive sulphur effect in power and distribution transformers failures and treatments. *Engineering Failure Analysis*, 2018, vol. 92, pp. 240–267. DOI: 10.1016/j.engfailanal.2018.05.018.
15. Arias Velásquez R.M., Mejía Lara J.V., Melgar A. Converting data into knowledge for preventing failures in power transformers. *Engineering Failure Analysis*, 2019, vol. 101, pp. 215–229. DOI: 10.1016/j.engfailanal.2019.03.027.
16. Tang W.H., Goulermas J.Y., Wu Q.H., Richardson Z.J., Fitch J. A probabilistic classifier for transformer dissolved gas analysis with a particle swarm optimizer. *IEEE Transactions on Power Delivery*, 2008, vol. 23, no. 2, pp. 751–759. DOI: 10.1109/TPWRD.2008.915812.
17. Yadaiah N., Ravi N. Internal fault detection techniques for power transformers. *Applied Soft Computing*, 2011, vol. 11, pp. 5259–5269. DOI: 10.1016/j.asoc.2011.05.034.
18. Duval M., De Pabla A. Interpretation of gas-in-oil analysis using new IEC publication 60599 and IEC TC 10 databases. *IEEE Electrical Insulation Magazine*, 2001, vol. 17, no. 2, pp. 31–41. DOI: 10.1109/57.917529.
19. Gopal S., Patro K., Sahu K.K. Normalization: a preprocessing stage. *International Advanced Research Journal in Science, Engineering and Technology*, 2015, vol. 2, no. 3, pp. 20–22. DOI: 10.17148/IARJSET.2015.2305.
20. Mueller A., Schmidt B., Keller C. Evaluating data preprocessing methods and their impact on machine learning performance. *Computational Intelligence (De Gruyter)*, 2022, vol. 20, no. 3, pp. 145–162. DOI: 10.1515/comp-2022-0278.
21. Wang L., Li X., Chen Y., Zhang Z. A review of modern data preprocessing techniques for machine learning. *2022 IEEE International Conference on Big Data*, 2022, pp. 120–125. DOI: 10.1109/BigData.2022.10714767.
22. Alsharif M., Abujarrad R., Al-Ramahi S. Data preprocessing techniques in Python using Pandas library. *2020 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2020, pp. 1–6. DOI: 10.1109/DSAA.2020.10575760.
23. Chicco D., Warrens M.J., Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 2021, vol. 7, article e623. DOI: 10.7717/peerj-cs.623.
24. Srivastava N., Agarwal M., Bhardwaj V., Lamba T. Comparative analysis of regression metrics for machine learning models. *Proceedings of the 2022 International Conference on Advances in Computing, Communication, and Intelligent Systems (ICACITE)*, 2022, pp. 475–480. DOI: 10.1109/ICACITE53722.2022.9823692.
25. Dewi A.L., Akrom M. Evaluating machine learning model performance using R², RMSE, and MAE. *Jurnal Informatika: Jurnal Pengembangan IT*, 2024, vol. 1, no. 2, pp. 45–53. DOI: 10.62411/jimat.v1i2.11053.
26. Kovrigin B.L., Lavrinovich A.V. Application of the determination coefficient to compare power transformer response waveforms obtained by nanosecond low-voltage impulse testing. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2018, vol. 329, no. 4, pp. 154–161. (In Russ.)
27. Maulud D.H., Abdulazeez A. A comparative review of linear regression and polynomial regression. *Journal of Applied Statistics and Theory*, 2020, vol. 15, no. 2, pp. 45–53. DOI: 10.38094/jastt1457.
28. Bartol K., Novak M., Horvat L., Zidar I. Linear regression as a baseline model for machine learning. *Sensors*, 2022, vol. 22, no. 5, article 188. DOI: 10.3390/s22051885.

29. Radhakrishnan A., Gupta P., Kumar S., Singh R. An overview of linear regression in machine learning. *IEEE Access*, 2022, vol. 10, pp. 15000–15008. DOI: 10.1109/ACCESS.2022.3141592.
30. Munkhdalai L., Enkhbayar T., Purev B., Davaadorj D. Improving linear regression interpretability with adaptive methods. *Mathematics*, 2020, vol. 8, no. 11, pp. 150–160. DOI: 10.1007/s11293-021-09711-8.
31. Bichri H., Chergui A., Hain M. Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 2024, vol. 15, no. 2, pp. 331–337. DOI: 10.14569/IJACSA.2024.0150235.
32. Xu Z., De A. Assessing model accuracy using random data split: a simulation study. *Journal of Biopharmaceutical Statistics*, 2022, vol. 33, pp. 131–139. DOI: 10.1080/10543406.2022.2089158.
33. Sáez J.A., Romero-Béjar J.L. Impact of k-fold cross-validation on regression model performance under dataset shift. *Mathematics*, 2022, vol. 10, no. 14, article 2538. DOI: 10.3390/math10142538.
34. Cai B., Pellegrini F., Pang M., de Moor C., Shen C., Charu V., Tian L. Bootstrapping the cross-validation estimate. *ISPRS Archives*, 2022, vol. XLVIII-4/W3-2022, pp. 111–118. DOI: 10.5194/isprs-archives-XLVIII-4-W3-2022-111-2022.
35. Nikitin N.O., Vychuzhanin P., Sarafanov M., Polonskaia I.S., Revin I., Barabanova I.V., Maximov G., Kalyuzhnaya A.V., Boukhanovsky A. Automated evolutionary approach for the design of composite machine learning pipelines. *Future Generation Computer Systems*, 2022, vol. 127, pp. 109–125. DOI: 10.1016/j.future.2021.08.022.
36. Cerrada M., Trujillo L., Hernández D., Correa Zevallos H.A., Macancela J.C., Román Cabrera D., Sánchez R.V. AutoML for feature selection and model tuning applied to fault severity diagnosis in spur gearboxes. *Mathematical and Computational Applications*, 2022, vol. 27, no. 1, article 6. DOI: 10.3390/mca27010006.
37. Osgonbaatar T., Matrenin P., Safaraliev M., Zicmane I., Rusina A., Kokin S. A rank analysis and ensemble machine learning model for load forecasting in the nodes of the central mongolian power system. *Inventions*, 2023. vol. 8. no. 5. pp. 114. DOI: 10.3390/inventions8050114.

Information about the authors

Vladislav A. Shelomentsev, Assistant, University of Tyumen, 6, Volodarsky street, Tyumen, 625003, Russian Federation; Postgraduate Student, Industrial University of Tyumen, 38, Volodarsky street, Tyumen, 625000, Russian Federation. shelomentsevva72@mail.ru <https://orcid.org/0000-0001-6672-9572>

Ilya S. Sukhachev, Cand. Sc., Associate Professor, Industrial University of Tyumen, 38, Volodarsky street, Tyumen, 625000, Russian Federation. suhachevis@tyuiu.ru <https://orcid.org/0000-0002-6220-6368>

Sergey V. Sidorov, Cand. Sc., Associate Professor, Industrial University of Tyumen, 38, Volodarsky street, Tyumen, 625000, Russian Federation. sidorovsv@tyuiu.ru <https://orcid.org/0009-0006-2246-3075>

Valery V. Sushkov, Dr. Sc., Professor, Nizhnevartovsk State University, 11, Dzerzhinsky street, Nizhnevartovsk, 628605, Russian Federation. sushkovvv50@mail.ru <https://orcid.org/0009-0005-9176-2426>

Rustam N. Khamitov, Dr. Sc., Professor, Tyumen Industrial University, 38, Volodarsky street, Tyumen, 625000, Russian Federation. apple_27@mail.ru; <https://orcid.org/0000-0001-9876-5471>

Petr V. Chepur, Cand. Sc., Associate Professor, Industrial University of Tyumen, 38, Volodarsky Street, Tyumen, 625000, Russian Federation. chepur_p_v@mail.ru <https://orcid.org/0000-0002-6722-469X>

Received: 19.03.2025

Revised: 07.04.2025

Accepted: 12.05.2025