

УДК 004.852, 004.67

ПРОГНОЗИРОВАНИЕ ПОТРЕБЛЕНИЯ ЭЛЕКТРИЧЕСКОЙ ЭНЕРГИИ ПРОМЫШЛЕННЫМ ПРЕДПРИЯТИЕМ С ПОМОЩЬЮ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Моргоева Анжелика Джабраиловна¹,
m.angelika-m@yandex.ru

Моргоев Ирбек Джабраилович¹,
m.irbek@yandex.ru

Клюев Роман Владимирович²,
kluev-roman@rambler.ru

Гаврина Оксана Александровна¹,
Gavrina-Oksana@yandex.ru

¹ Северо-Кавказский горно-металлургический институт (Государственный технологический университет),
Россия, 362011, г. Владикавказ, ул. Николаева, 44.

² Московский политехнический университет,
Россия, 107023, г. Москва, ул. Б. Семеновская, 38.

Актуальность исследования заключается в необходимости развития энергосберегающих подходов за счет применения средств интеллектуального анализа данных для повышения эффективности процесса принятия управленческих решений и, как следствие, более оптимального использования энергетических ресурсов. В частности, прогнозирование потребления электрической энергии промышленного объекта упростит процесс принятия управленческих решений и будет способствовать минимизации затрат электроэнергии на производство единицы продукции. Наличие точного прогноза позволит задействовать резервные мощности в часы пиковой нагрузки на электроэнергетический комплекс. На практике для реализации прогнозирования не всегда подходят существующие методы расчета нагрузки на электросеть, поэтому исследование носит междисциплинарный характер, сочетая в себе важную практическую значимость и выработку новых рекомендаций в части применения алгоритмов машинного обучения.

Целью исследования является анализ научных работ, содержащих предложения по повышению точности определения энергетических нагрузок с помощью интеллектуального анализа данных, а также разработка модели машинного обучения, позволяющей создавать достоверный прогноз потребления электроэнергии для промышленного предприятия.

Объекты: промышленное предприятие, которое характеризуется сложностью определения энергетических характеристик технологического оборудования.

Методы: аналитический метод, методы математической статистики, методы машинного обучения, комплексного обобщения научных достижений и практического опыта применения средств обработки данных в задачах прогнозирования нагрузки на электросеть.

Результаты. Проведен обзор литературных источников, освещающих вопросы применения интеллектуального анализа данных в управлении энергопотреблением, и представлены основные результаты прогнозирования суммарных объемов потребления электроэнергии по данным промышленного объекта. Рассмотрены методы интеллектуального анализа данных, применяемые для решения задач энергосбережения для различных объектов. Построена модель машинного обучения на основе алгоритма градиентного бустинга библиотеки CatBoost, позволяющая получить прогноз потребления электроэнергии по месяцам с уровнем надежности 92 %. Результаты проведенного исследования актуальны для принятия решений на тактическом и стратегическом уровнях управления предприятием для среднесрочного (помесячного) и долгосрочного (от года до нескольких лет) прогнозирования электрических нагрузок соответственно.

Ключевые слова:

Ресурсосберегающие технологии, энергосбережение, прогнозирование, интеллектуальный анализ данных, машинное обучение, градиентный бустинг.

Введение

Оптимальное управление ресурсами является важной задачей менеджмента предприятия. В понятие термина «ресурсосберегающие технологии» заложено обеспечение минимально возможных затрат ресурсов на производство единицы продукции при сохранении заданного уровня качества. В данной работе рассматривается электроэнергия как важный ресурс, необходимый для функционирования не только промышленных предприятий, но и для большинства других объектов гражданского назначения.

Одним из последствий роста глобального использования вычислительной и бытовой техники является увеличение потребления электрической энергии, что влечет за собой ряд экологических проблем. Отдельно стоит отметить, что вопросы энергосбережения актуальны и с точки зрения экономических выгод, получаемых при рациональном управлении энергетическими ресурсами. При комплексном рассмотрении причин возрастания электрических нагрузок в контексте вычислительной техники и средств обработки данных можно заметить следующие особенности.

С одной стороны, развитие и применение методов интеллектуального анализа данных, в частности для задачи прогнозирования электрических нагрузок, способствует уменьшению и рационализации использования ресурсов. Однако, с другой стороны, для выполнения процедур машинного обучения требуются значительные вычислительные ресурсы и, следовательно, увеличение потребления электрической энергии.

В данной работе приведены результаты исследований по поиску возможных путей сокращения электрических нагрузок. В частности, приведены примеры эффективных решений в части сокращения расходов на электроэнергию с помощью применения интеллектуальной обработки данных, и проведено прогнозирование электрических нагрузок с целью минимизации затрат на электроэнергию за счет наличия точного прогноза и дальнейшего учета прогнозных данных при планировании расходов на электроэнергию.

Обзор литературы

Исследованиям и разработке энергосберегающих технологий посвящены научные работы в России и за рубежом [1–14].

Поэтому первым этапом исследования стал аналитический обзор литературных источников по теме интеллектуального анализа данных, применяемого для решения задач энергосбережения в различных сферах деятельности. Рассмотрим некоторые работы более подробно.

В [15] учеными из Токио представлен подход, позволяющий оптимально управлять энергопотреблением здания (на примере гостиницы). Метод представляет собой комбинацию методов машинного обучения и метаэвристики, и его разработчики утверждают, что применение этого подхода позволило сократить операционные расходы на потребление электроэнергии более чем на 10 %.

Согласно [15] метод машинного обучения на основе глубоких нейронных сетей использовался для прогнозирования оптимальной работы интегрированных систем градирен, а метод метаэвристики – для оптимизации работы других компонентов энергетического комплекса. Исследователи поясняют, что создание гибридного метода позволяет учитывать нелинейность данных электропотребления для некоторых видов оборудования.

Ученые в [15] дифференцируют методы оптимизации энергосистемы на две основные группы. Первая группа связана с использованием мощных решений, таких как смешанно-целочисленное линейное программирование. В другую группу исследователи отнесли метаэвристические методы и модели машинного обучения, применяемые для учета нелинейности данных. Ученые концентрируют внимание на второй группе методов, так как линейные модели позволяют моделировать лишь ограниченные случаи, основаны на сравнении сценариев и сопряжены со значительными затратами времени и вычислительных ресурсов.

В результате исследования, описанного в [15], с помощью метаэвристики были определены оптимальные графики работы каждого энергопотребляющего устройства на 24 часа. Чтобы снизить вычислительные нагрузки, в метаэвристику была интегрирована регрессионная модель. Как утверждают авторы, в работе были применены модели линейного множественного регрессионного анализа, модель случайного леса и глубокая нейронная сеть (Deep Neural Network – DNN). Параметры модели случайного леса были установлены на 100 деревьев поиска, для DNN были заданы три скрытых слоя (количество узлов: 100, 100, 500). В связи с тем, что в [15] предполагалось оптимизировать потребление электроэнергии, кондиционирования и горячего водоснабжения одновременно, учеными была поставлена и решена задача сложной нелинейно оптимизации. При этом с помощью методов машинного обучения и метаэвристик были получены оптимальные графики работы устройств, что позволило сократить суточные эксплуатационные расходы. Однако, как утверждают разработчики, данный гибридный метод еще не до конца изучен, и необходимо дальнейшее его тестирование на других объектах.

Российскими учеными из Новосибирского государственного технического университета (НГТУ) совместно с компанией, занимающейся цифровизацией энергетики, было проведено прогнозирование энергопотребления крупных энергоемких промышленных предприятий [16]. Исследователи ставили задачу – автоматизировать работу экспертов, занимающихся составлением почасового прогноза потребления электрической энергии на сутки. Использованные методы: за основу была взята модель рекуррентных нейронных сетей с долгосрочной памятью (Long Short-Term Memory Recurrent Neural Networks), разработанная в НГТУ. Рассматривались следующие предприятия: нефтедобывающая и нефтеперерабатывающая компания, мебельный завод, завод по производству стали и труб, компания по добыче и обработке драгоценных камней, горно-обогатительный комбинат, металлургический завод. Исходными данными были: фактическое потребление электроэнергии, данные ремонта и метеорологические данные из открытых источников. Стоит отметить, что полученные с помощью моделей машинного обучения прогнозы сравнивались не только с фактическими данными потребления электроэнергии, но и с прогнозными значениями экспертов. Экономическая эффективность рассчитывалась от снижения ошибки на 1 %. Исследователи констатируют, что в целом в результате экспериментов получена незначительная разница между точностью прогнозов экспертов и нейронной сети. Исключение составила компания по добыче и обработке драгоценных камней, где снижение ошибки, полученное в результате использования модели машинного обучения, составило от 1 до 2,7 %.

В работе [17] представлены результаты применения машинного обучения для прогнозирования потребления энергии охлаждения в офисных зданиях с учетом поведения людей. Исследователи отмечают,

что прогнозирование энергопотребления в зданиях играет ключевую роль в принятии решений по энергоэффективности. С развитием аналитики данных в последние годы был разработан ряд моделей прогнозирования энергопотребления зданий на основе машинного обучения, однако существующие модели прогнозирования недостаточно учитывают поведение людей, находящихся в зданиях. Чтобы устранить этот пробел, в [17] представлен подход машинного обучения для прогнозирования энергопотребления здания с учетом поведения людей. При разработке модели были протестированы и сопоставлены четыре алгоритма машинного обучения с точки зрения точности прогнозирования и вычислительной эффективности: деревья классификации и регрессии (Classification and Regression Trees – CART), деревья ансамблевого бэггинга (Ensemble Bagging Trees – EBT), искусственные нейронные сети (Artificial Neural Networks – ANN) и глубокие нейронные сети (DNN). Результаты моделирования продемонстрировали большое влияние переменных, рассмотренных в исследовании, на целевой результат – потребление электроэнергии [17]. Деятельность работников офисного здания повлияла на потребление энергии более чем в 7 раз. Столь высокая производительность показывает потенциал предлагаемого подхода. Ученые полагают, что такой подход может способствовать лучшему пониманию влияния поведения людей на потребление электроэнергии в здании и определить возможности для поведенческих мер по энергосбережению.

Работа [18] посвящена прогнозированию энергопотребления и оценке влияющих факторов с использованием подхода машинного обучения на полномасштабных очистных сооружениях. Поскольку очищение сточных вод является энергоемким процессом и вносит основной вклад в затраты на очистные сооружения, анализ и прогнозирование энергопотребления имеют важное значение при проектировании и эксплуатации очистных сооружений. В [18] было проведено исследование влияния на суточное потребление электроэнергии очистными сооружениями следующих факторов: сточных вод, гидравлических и климатических параметров. Данные энергопотребления были взяты из архива комплекса очистных сооружений, расположенного в Мельбурне, за шесть лет (2014–2019 гг.). Учеными были применены методы инженерии данных для объединения функций из разных ресурсов. С этой целью были использованы четыре различных алгоритма выбора функций (Feature Selection – FS), чтобы выявить отношения между этими переменными и выбрать наиболее релевантные переменные для обучения моделей машинного обучения (Machine Learning – ML). Далее были применены искусственные нейронные сети (ANN). Два алгоритма деревьев решений – Gradient Boosting Machine (GBM) и Random Forest (RF) – были изучены для прогнозирования потребления электроэнергии с оценкой доверительного интервала 95 %. Результаты исследованных в [18] алгоритмов показали, что общий азот, химическая потребность в кислороде имеют наибольшее влияние на потребление энергии

очистными сооружениями. Более того, у GBM был лучший прогноз производительности среди всех других алгоритмов регрессии, а доверительный интервал, равный 95 %, показал разумную полосу ошибок прогноза.

Немецкие ученые решили другую актуальную задачу, заключающуюся в разработке метода анализа кривых электроэнергетической нагрузки для обнаружения аномалий. В [19] исследователи утверждают, что точное понимание кривых энергетической нагрузки является ключом к эффективному управлению энергосистемами предприятия и основой для нескольких энергетических приложений (например, прогнозов, обнаружения аномалий). Также ученые отмечают, что анализ кривой нагрузки является темой исследования, имеющей практическое значение во многих областях. Кроме того, исследователи акцентируют внимание на отсутствии методов для оценки различных временных переходов между энергетическими состояниями при анализе кривых нагрузок. В [19] представлена новая методология, сочетающая методы одномерной кластеризации без учителя и методы, основанные на многомерном прогнозировании, основанном на использовании искусственной нейронной сети. Для обнаружения аномалий в управлении энергоэффективностью эти методы применены и проверены на реальных заводских данных. Кроме того, ученые утверждают, что несмотря на то, что демонстрация методологии сосредоточена на анализе кривых электрической нагрузки, представленный подход может быть перенесен на другие формы энергии (например, сжатый воздух, производственный газ и т. д.).

Проанализировав современные исследования в области применения интеллектуального анализа данных для получения энергосберегающих решений, был сделан вывод, подтверждающий актуальность подобных разработок. Стоит заметить, что в большинстве из предлагаемых в литературе методов использованы комбинации различных алгоритмов машинного обучения и эвристических подходов. С учетом того, что исходные данные к исследованию являются данными об электропотреблении промышленного предприятия, которое характеризуется сложностью определения энергетических характеристик технологического оборудования, было принято решение провести прикладное исследование.

Прогнозирование потребления электрической энергии

На начальном этапе исследования был проведен сбор информации о предприятии и исходных данных по его электропотреблению. С учетом соглашения о неразглашении (non-disclosure agreement – NDA) наименование предприятия обезличено. Его краткая характеристика следующая: небольшое промышленное предприятие, технологическое оборудование которого энергоемкое и устаревшее, выпуск продукции неравномерный в течение года. Учет потребления электроэнергии ведется дифференцировано для потребления на технические нужды и освещение. Предприятие имеет 14 подразделений, в которых 2 не

имеют расходов на технические нужды (административные корпуса).

В качестве исходных данных были взяты: фактические данные помесичного потребления электроэнергии за период с 2015 по 2021 гг., метеорологические данные из открытых источников информации.

Исследование было выполнено на языке программирования Python (v. 3.10.0) в среде программирования Jupyter Notebook. Были использованы библиотеки Numpy и Pandas (для вычислений и манипуляций с

данными), Matplotlib и Seaborn (для визуализации данных), Scikit-learn, XGBoost и CatBoost (для предобработки данных и загрузки экземпляров моделей регрессии).

На первом этапе исследования был проведен анализ исходного набора данных (дата-сета). На рис. 1 приведены графики потребления электроэнергии по годам по суммарному потреблению электроэнергии, потреблению на технические нужды и на освещение соответственно.

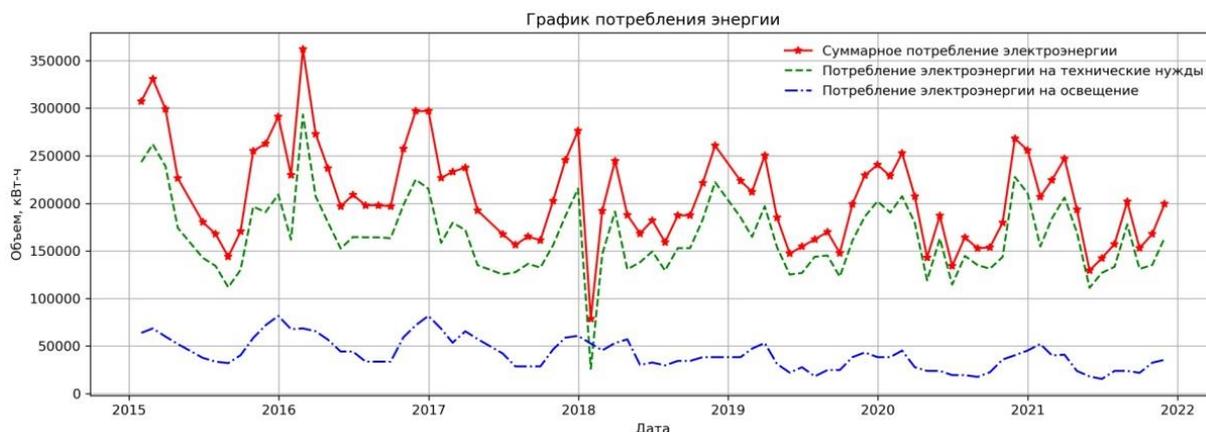


Рис. 1. График потребления электроэнергии предприятием за 2015–2021 гг.

Fig. 1. Schedule of electricity consumption by the enterprise for 2015–2021

Для более детального исследования характера как суммарного потребления электроэнергии, так и на технические нужды и освещение за рассматриваемый период, были построены графики, изображенные на

рис. 2–4, отражающие для большей наглядности помимо фактических значений потребления еще и кривые средних значений расхода электроэнергии за весь период.

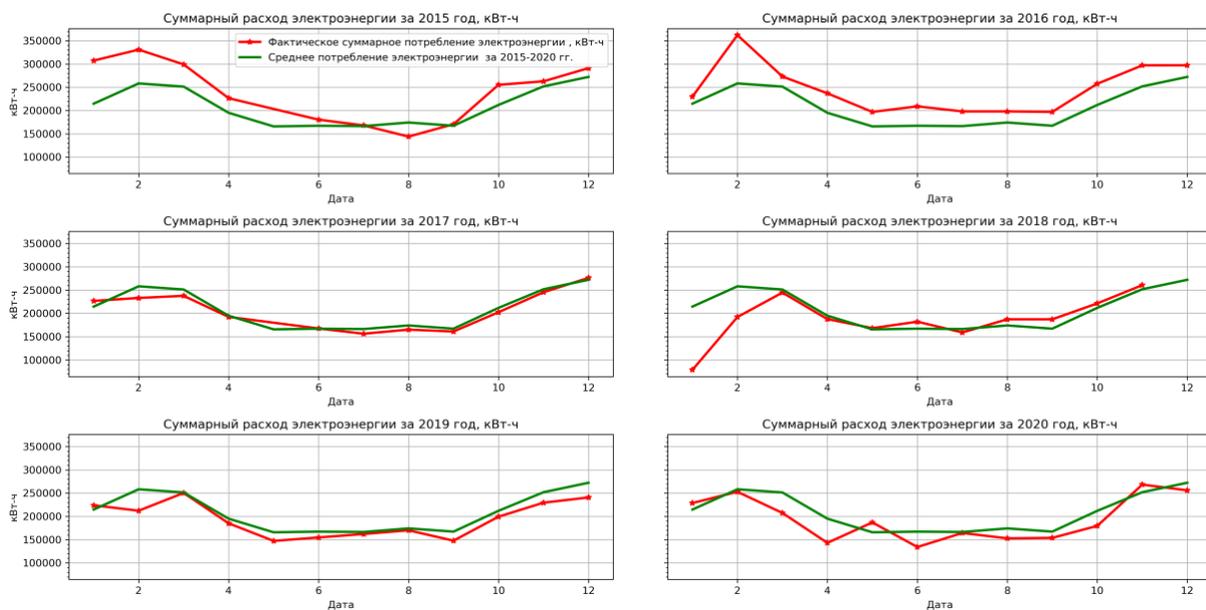


Рис. 2. Графики суммарного потребления электроэнергии по годам

Fig. 2. Schedules of total electricity consumption by years

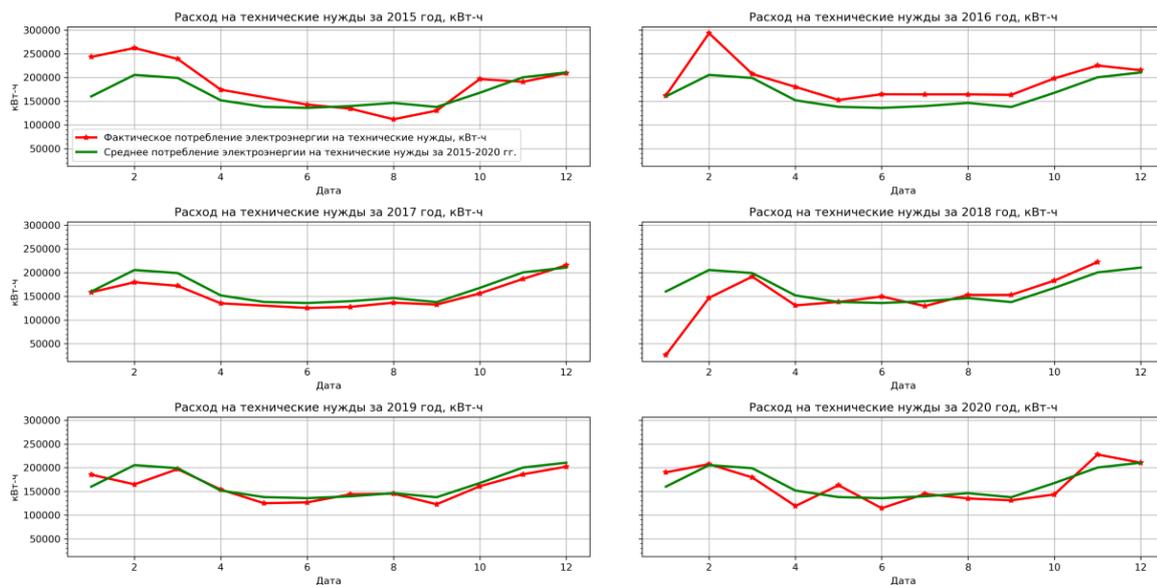


Рис. 3. Графики потребления электроэнергии на технические нужды по годам

Fig. 3. Schedules of electricity consumption for technical needs by years

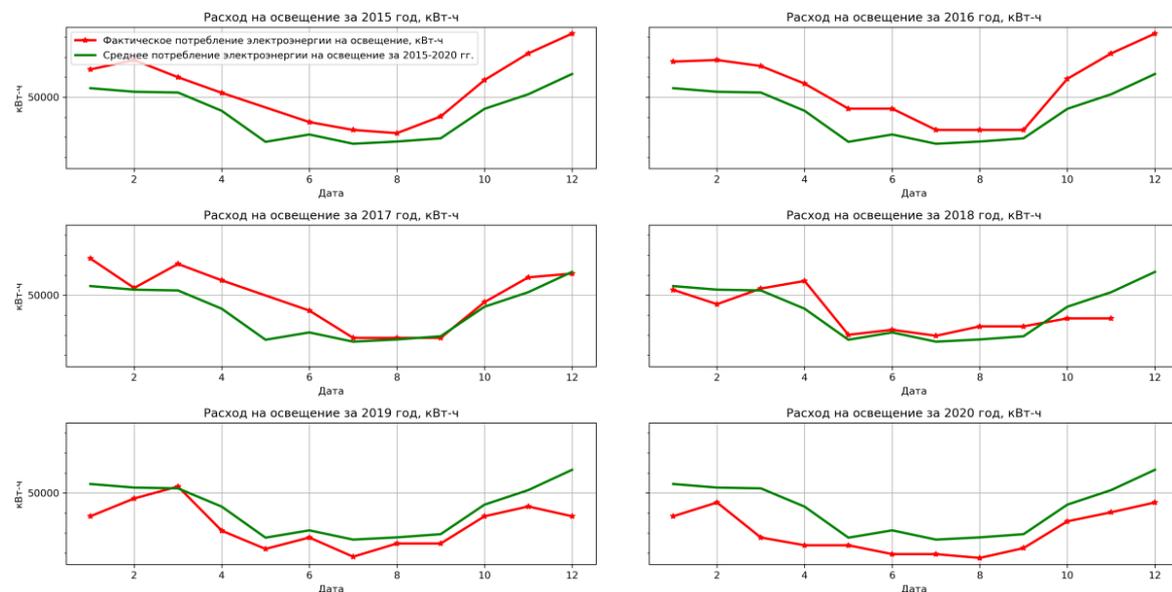


Рис. 4. Графики потребления электроэнергии на освещение по годам

Fig. 4. Schedules of electricity consumption for lighting by years

Следующим этапом стала предобработка данных, определение признаков и целевой переменной.

Временные данные были разбиты на такие признаки, как: месяц, год, количество дней в месяце, день недели. Так как прогнозирование проводилось по месяцам, для каждого месяца были добавлены признаки фактического расхода электроэнергии за последние четыре года, а также за предшествующие этому месяцу 4 месяца. Кроме того, дата-сет был дополнен погодными данными: температура окружающей среды, влажность, температура точки росы, скорость ветра. Для каждого структурного подразделения промышленного объекта был рассчитан фактор, характеризующий удельный вес потребления электроэнергии соответствующим объектом как отношение потребляемой им мощности к суммарному потреблению пред-

приятия за месяц. В табл. 1 приведены описания признаков и соответствующие им обозначения. Целевой переменной выбран показатель суммарного потребления электроэнергии. На рис. 5 представлены диаграммы, отражающие удельные веса потребления электроэнергии каждого подразделения. Как можно заметить по диаграммам, наибольший вклад в суммарное потребление и потребление электроэнергии на технические нужды вносит подразделение W_D_1, являющееся главным технологическим участком. Стоит отметить, что в диаграмме весов потребления на технические нужды нет участка W_D_12, но этот участок является главным потребителем электроэнергии на освещение. Это объяснимо тем, что данное подразделение представляет собой административный корпус.

Таблица 1. Описание признаков

Table 1. Features description

Название признака Feature name	Обозначение Designation
Целевая переменная (y)/Target variable (y)	SUM_consumption
Месяц/Month	month
Год/Year	year
День недели/Day of week	day-of-week
Количество дней в месяце Amount of days in a month	days_in_month
Фактическое потребление электроэнергии 1–4 года назад Real electricity consumption 1–4 years ago	1_year_ago 2_years_ago 3_years_ago 4_years_ago
Фактическое потребление электроэнергии 1–4 месяца назад Real electricity consumption 1–4 months ago	1_month_ago 2_months_ago 3_months_ago 4_months_ago
Удельный вес первого–четырнадцатого подразделения в суммарное потребление электроэнергии Specific weight of the first–fourteenth division in the total electricity consumption	W_D_1...W_D_14
Температура окружающей среды (среднее значение за месяц) Ambient temperature (monthly average)	T
Скорость ветра (среднее значение за месяц) Wind speed (monthly average)	Ff
Влажность (среднее значение за месяц) Humidity (monthly average)	U
Температура точки росы (среднее значение за месяц) Dew point temperature (monthly average)	Td

Для определения значимости признаков была построена тепловая карта – матрица корреляции, рассчитанная по (1), где x и y – попарно перебираемые признаки (рис. 6). В результате анализа матрицы корреляции было принято решение оставить все признаки в модели.

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma(x) \cdot \sigma(y)}, \quad (1)$$

Далее дата-сет был разбит на тренировочную и тестовую выборки в отношении 80:20. Для построения моделей были использованы алгоритмы машинного обучения, представленные в табл. 2. Это модели линейной регрессии (LinearRegression, Lasso, Ridge), k -ближайших соседей (KNeighborsRegressor), дерева решений (DecisionTreeRegressor), случайного леса (RandomForestRegressor), линейной регрессии опорных векторов (LinearSVR), многослойного перцептрона (MLPRegressor), экстремального градиентного бустинга (XGBoost) и градиентного бустинга, основанная на деревьях решений, (CatBoost). Подбор параметров моделей происходил с помощью инструмента перекрестной проверки GridSearchCV. Метриками качества были выбраны средняя абсолютная ошибка прогноза в процентах (mean absolute percentage error – MAPE), рассчитываемая по (2) и показывающая в процентном соотношении уровень ошибок, допускаемых алгоритмом, и коэффициент детерминации (R^2 score), вычисляемый по (3) и характеризующий точность алгоритма регрессии.

$$MAPE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100 \%, \quad (2)$$

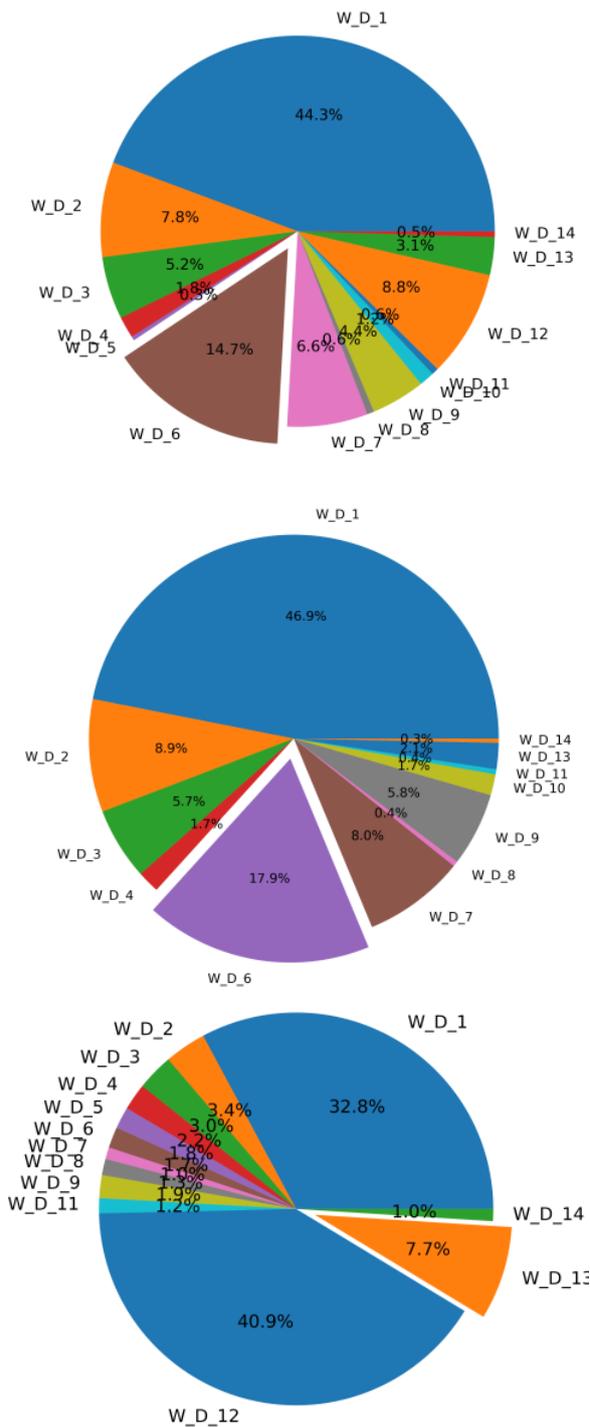


Рис. 5. Диаграммы удельных весов подразделений для суммарного потребления электроэнергии, на технические нужды и освещение

Fig. 5. Diagrams of unit specific weights for total electricity consumption, for technical needs and lighting

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (3)$$

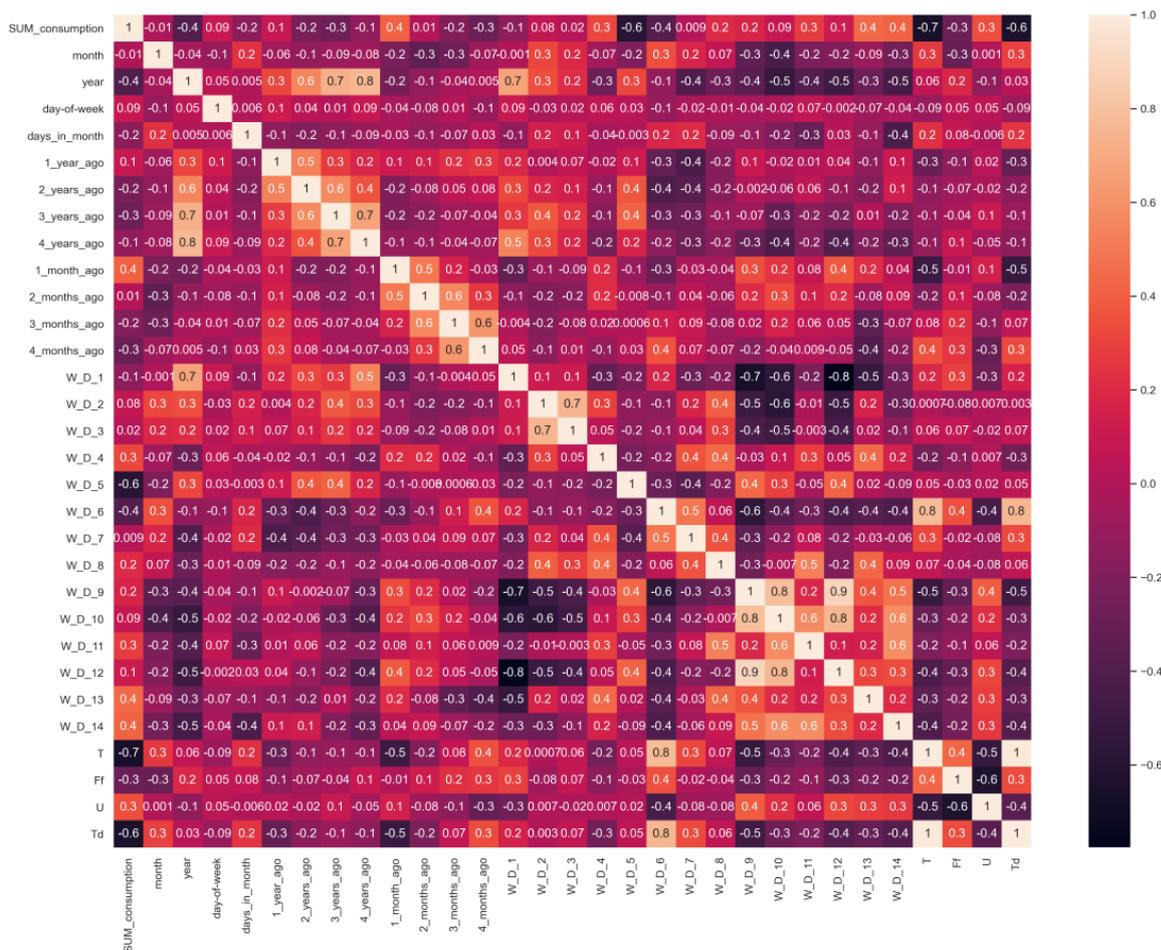


Рис. 6. Матрица корреляции признаков и целевой переменной
 Fig. 6. Correlation matrix of features and target variable

Таблица 2. Результаты метрик качества моделей
 Table 2. Results of model quality metrics

Модель Model	Метрики качества прогноза/Forecast quality metrics					
	Прогнозирование по суммарным нагрузкам Forecasting by total loads		Прогнозирование по потреблению на технические нужды Forecasting of consumption for technical needs		Прогнозирование по потреблению на освещение Forecasting of lighting consumption	
	MAPE	R ² score	MAPE	R ² score	MAPE	R ² score
Линейная регрессия Linear Regression	16,09	0,3	89,98	-2,48	20,35	0,19
Лассо регрессия Lasso Regression	16,55	0,23	90,94	-2,55	20,27	0,19
Ридж регрессия Ridge Regression	12,23	0,39	67,47	-0,81	19,95	0,33
К-ближайших соседей KNeighborsRegressor	11,85	0,39	45,33	0,11	13,9	0,73
Дерево решений DecisionTreeRegressor	9,89	0,69	38,53	0,38	9,25	0,56
Случайный лес Random Forest Regressor	13,11	0,46	46,86	0,11	8,22	0,83
Линейная регрессия опорных векторов LinearSVR	18,51	-0,36	65,33	-0,73	39,19	-0,45
Многослойный перцептрон MLPRegressor	12,71	0,37	100	-10,79	27,1	-0,58
Экстремальный градиентный бустинг XGBRegressor	9,97	0,68	34,98	0,43	7,82	0,75
Градиентный бустинг CatBoostRegressor	7,95	0,77	39,86	0,33	6,98	0,94

Было проведено прогнозирование по нескольким сценариям: отдельно по фактическому потреблению на технические нужды, отдельно по фактическому потреблению на освещение и прогнозирование по суммарным данным потребления электроэнергии. Как следует из табл. 2, в случае прогнозирования по данным электропотребления на технические нужды наилучший результат показала модель, основанная на алгоритме регрессии экстремального градиентного бустинга XGBoost, ошибка прогноза составила 34,98 %. При прогнозировании расхода электроэнергии на освещение наиболее точный прогноз дала модель, базирующаяся на алгоритме градиентного бустинга CatBoost, ошибка прогноза которой составила 6,98 % при доле объясненной дисперсии 0,94, что является очень точным результатом прогноза. Стоит отметить, что моделирование по данным потребления электроэнергии на освещение является более точным, чем по данным потребления электроэнергии на технические нужды, что объясняется меньшим разбросом исходных данных относительно среднего значения.

В случае же прогнозирования по суммарным данным наилучший результат показала модель, основанная на алгоритме градиентного бустинга CatBoost

[20]. Ошибка прогноза данной модели составила 7,95 %, что на 1,94 меньше по сравнению с моделью дерева решений Decision Tree Regressor и на 2,02 меньше, чем ошибка прогноза модели экстремального градиентного бустинга XGBoost. Таким образом, модель, основанная на алгоритме CatBoost, дает прогноз с достоверностью 92,05 % при доле объяснения признаками целевого результата в 0,77, что позволяет классифицировать построенную в результате исследования модель машинного обучения как довольно точную. Графики спрогнозированных моделью CatBoost значений и фактических данных приведены на рис. 7. CatBoost является библиотекой с открытым исходным кодом, использующей «небрежные» деревья для построения сбалансированного дерева решений, что способствует более эффективному обучению. Особенностью данного алгоритма является простота при работе с категориальными переменными. Разработчик данной библиотеки – компания Яндекс. Стоит отметить, что в классе алгоритмов градиентного бустинга библиотека CatBoost с момента появления заняла лидирующие позиции и является достойным конкурентом библиотеки экстремального градиентного бустинга XGBoost, что и подтверждает наше исследование.

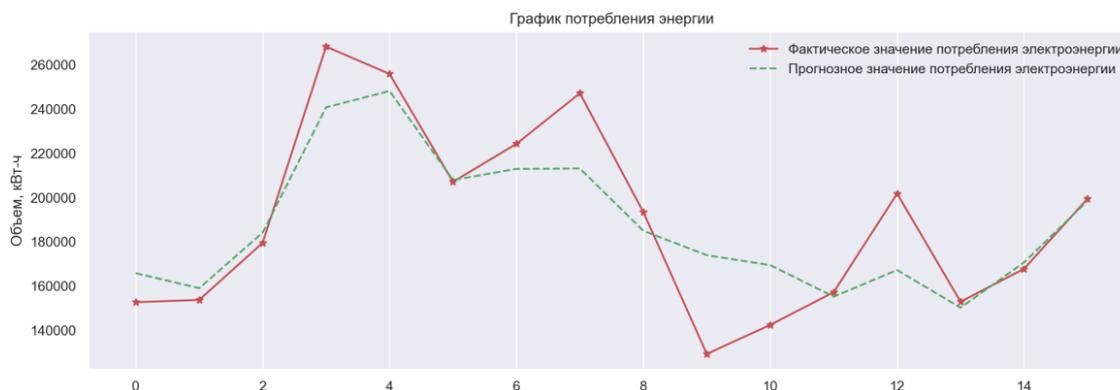


Рис. 7. График фактических и прогнозных значений целевой переменной

Fig. 7. Schedule of real and forecast values of the target variable

Кроме того, был проведен эксперимент сравнения результатов прогноза по данным суммарного потребления электроэнергии и суммы результатов прогнозов по частям (технические нужды и освещение). Однако сумма результатов прогнозов по частям значительно отличалась от проверочного набора данных, что подтверждает необходимость проведения прогноза по данным суммарного потребления электроэнергии.

Заключение

В результате проведенного исследования была получена модель машинного обучения, основанная на алгоритме градиентного бустинга библиотеки CatBoost. Точность прогноза потребления электрической энергии по данной модели составила более 92 %, что позволило классифицировать прогнозные значения как достоверные и считать полученную модель адекватной экспериментальным данным. Кроме того, были сделаны следующие выводы.

Во-первых, задача является актуальной как для научного сообщества с точки зрения исследования применяемых методов прогнозирования, так и для предприятий с точки зрения экономических выгод от внедрения энергосберегающих систем.

Во-вторых, несмотря на то, что электроэнергия является одним из ключевых ресурсов для производства и в сфере услуг, прогнозирование потребности для каждой из областей использования электроэнергии довольно специфично, требует учета множества факторов и их интеллектуального анализа.

В-третьих, для точного прогнозирования необходимы данные о технологическом оборудовании, его мощности, также необходим учет более детальных данных о всех потребителях электроэнергии рассматриваемого объекта, объемах выпускаемой продукции и других данных, влияющих на потребление электроэнергии.

В-четвертых, интеллектуальная обработка данных полностью не заменяет работу экспертов в оценке прогнозирования электропотребления, что связано с учетом экспертами дополнительных факторов, в том числе случайных, и сложностью интерпретации результатов, полученных с помощью моделей машин-

ного обучения. Поэтому необходим комплексный подход к процессу расчета потребления электрической энергии и проведение новых исследований с целью поиска наиболее точных результатов прогнозирования.

СПИСОК ЛИТЕРАТУРЫ

1. Статистический анализ повреждений в карьерной сети горно-металлургического комбината / Р.В. Клюев, И.И. Босиков, О.А. Гаврина, К.С. Крысанов // Известия Тульского государственного университета. Науки о Земле. – 2021. – № 1. – С. 168–178.
2. Луковенко А.С. Повышение надежности и качества электрооборудования потребителей тяговых подстанций переменного тока: автореф. дис. ... канд. техн. наук. – Красноярск, 2016. – 21 с.
3. Прогнозирование удельного потребления электроэнергии обогатительной фабрики / Р.В. Клюев, О.А. Гаврина, В.Н. Хетагуров, С.Г. Зассеев, Б.З. Умиров // Горный информационно-аналитический бюллетень (научно-технический журнал). – 2020. – № 11-1. – С. 135–145.
4. Анализ потерь электроэнергии в системе электроснабжения обогатительной фабрики / Р.В. Клюев, В.И. Голик, И.И. Босиков, О.А. Гаврина // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2021. – Т. 332. – № 10. – С. 7–16.
5. Energy saving in metro ventilation system based on multi-factor analysis and air characteristics of piston vent / Deqiang He, Xiaoliang Teng, Yanjun Chen, Bin Liu, Heliang Wang, Xianwang Li, Rui Ma // Applied Energy. – 2022. – V. 307. – 118295.
6. Комплексный анализ применения эффективных технологий для повышения устойчивого развития природно-технической системы / Р.В. Клюев, И.И. Босиков, А.В. Майер, О.А. Гаврина // Устойчивое развитие горных территорий. – 2020. – № 2. – С. 283–290.
7. Da-sheng Lee, Yan-Tang Chen, Shih-Lung Chao Universal workflow of artificial intelligence for energy saving // Energy Reports. – 2022. – V. 8. – P. 1602–1633.
8. Energy-saving potential prediction models for large-scale building: a state-of-the-art review / Xiu'e Yang, Shuli Liu, Yuliang Zou, Wenjie Ji, Qunli Zhang, Abdullahi Ahmed, Xiaojing Han, Yongliang Shen, Shaoliang Zhang // Renewable and Sustainable Energy Reviews. – 2022. – V. 156. – 111992.
9. Energy saving technologies and mass-thermal network optimization for decarbonized iron and steel industry: A review / R.Q. Wang, L. Jiang, Y.D. Wang, A.P. Roskilly // Journal of Cleaner Production. – 2020. – V. 274. – 122997.
10. Tiancheng Shang, Peihong Liu, Junxiong Guo How to allocate energy-saving benefit for guaranteed savings EPC projects? A case of China // Energy. – 2020. – V. 191. – 116499.
11. Energy-saving effect of integrated cooling unit with rotary booster and compressor for data center / Yu Liu, Guoyuan Ma, Lianzheng Xue, Feng Zhou, Lei Wang // International Journal of Refrigeration. – 2020. – V. 119. – P. 366–375.
12. Multi-dimensional analysis of air-conditioning energy use for energy-saving management in university teaching buildings / Xinyue Li, Shuqin Chen, Hongliang Li, Yunxiao Lou, Jiahe Li // Building and Environment. – 2020. – V. 185. – 107246.
13. Estimation of energy consumption in machine learning / E. Garcia-Martin, C. Faviola Rodrigues, G. Riley, H. Grahn // Journal of Parallel and Distributed Computing. – 2019. – V. 134. – P. 75–88.
14. Ghoddsi H., Creamer G.G., Rafizadeh N. Machine learning in energy economics and finance: a review // Energy Economics. – 2019. – V. 81. – P. 709–727.
15. Shintaro Ikeda, Tatsuo Nagai. A novel optimization method combining metaheuristics and machine learning for daily optimal operations in building energy and storage systems // Applied Energy. – 2021. – V. 289. – 116716.
16. Прогнозирование и планирование потребления электроэнергии с помощью machine learning (эксперимент). URL: <https://habr.com/ru/post/577732/> (дата обращения 10.01.2022).
17. Amasyalia K., El-Gohary N. Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings // Renewable and Sustainable Energy Reviews. – 2021. – V. 142. – 110714.
18. Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach / Faramarz Bagherzadeh, Amirreza Shojaei Nouri, Mohamad-Javad Mehrani, Suresh Thennadil // Process Safety and Environmental Protection. – 2021. – V. 154. – P. 458–466.
19. Machine learning based analysis of factory energy load curves with focus on transition times for anomaly detection / D. Flick, C. Keck, Ch. Herrmann, S. Thiede // Procedia CIRP. – 2020. – V. 93. – P. 461–466.
20. CatBoost Regressor. URL: https://catboost.ai/en/docs/concepts/python-reference_catboostregressor (дата обращения 11.01.2022).

Поступила 17.06.2022 г.

Информация об авторах

Моргоева А.Д., аспирант Северо-Кавказского горно-металлургического института (Государственного технологического университета).

Моргоев И.Д., аспирант Северо-Кавказского горно-металлургического института (Государственного технологического университета).

Клюев Р.В., доктор технических наук, профессор кафедры техники низких температур им. П.Л. Капицы Московского политехнического университета.

Гаврина О.А., кандидат технических наук, доцент кафедры электроснабжения промышленных предприятий Северо-Кавказского горно-металлургического института (государственного технологического университета).

UDC 004.852, 004.67

FORECASTING OF ELECTRIC ENERGY CONSUMPTION BY AN INDUSTRIAL ENTERPRISE USING MACHINE LEARNING METHODS

Anzhelika D. Morgoeva¹,
m.angelika-m@yandex.ru

Irbek D. Morgoev¹,
m.irbek@yandex.ru

Roman V. Klyuev²,
kluev-roman@rambler.ru

Oksana A. Gavrina¹,
Gavrina-Oksana@yandex.ru

¹ North Caucasian Institute of Mining and Metallurgy (State Technological University),
44, Nikolaev street, Vladikavkaz, 362011, Russia.

² Moscow Polytechnic University,
38, B. Semenovskaya street, Moscow, 107023, Russia.

The relevance of the research is caused by the need to develop energy-saving approaches through the use of data mining tools to improve the efficiency of the management decision-making process and, as a result, more optimal use of energy resources. In particular, forecasting the consumption of electric energy of an industrial facility will simplify the process of making managerial decisions and will help minimize the cost of electricity for the production of a unit of production. The availability of an accurate forecast will enable the use of reserve capacities during peak load hours for the electric power complex. In practice, the existing methods of calculating the load on the power grid are not always suitable for forecasting, so the study is interdisciplinary in nature, combining important practical significance and the development of new recommendations regarding the use of machine learning algorithms.

The main aim of the research is to analyze scientific papers containing proposals to improve the accuracy of determining energy loads using data mining, as well as to develop a machine learning model that allows you to create a reliable forecast of electricity consumption for an industrial enterprise.

Objects of the research is an industrial enterprise that is characterized by the complexity of determining the energy characteristics of technological equipment.

Methods: analytical method, methods of mathematical statistics, methods of machine learning, complex generalization of scientific achievements and practical experience in the use of data processing tools in the tasks of predicting the load on the power grid.

Results. A review of literature sources covering the application of data mining in energy consumption management is carried out, and the main results of forecasting total electricity consumption according to industrial facility data are presented. The methods of data mining used to solve energy saving problems for various objects are considered. A machine learning model is built based on the gradient boosting algorithm of the CatBoost library, which allows obtaining a forecast of electricity consumption by months with a reliability level of 92 %. The results of the study are relevant for decision-making at the tactical and strategic levels of enterprise management for medium-term (monthly) and long-term (from a year to several years) forecasting of electrical loads, respectively.

Key words:

Resource-saving technologies, energy saving, forecasting, data mining, machine learning, gradient boosting.

REFERENCES

- Klyuev R.V., Bosikov I.I., Gavrina O.A., Krysanov K.S. Statistical analysis of damages in the opening network of the mining and metallurgical combine. *Bulletin of the Tula State University. Earthsciences*, 2021, vol. 1, pp. 168–178. In Rus.
- Lykovenko A.S. *Povyshenie nadezhnosti i kachestva elektrosnabzheniya potrebiteley tyagovykh podstantsiy peremennogo toka*. Avtoreferat Dis. Kand. nauk [Improving the reliability and quality of power supply to consumers of traction substations of alternating current. Cand. Diss. Abstract]. Krasnoyarsk, 2016. 21 p.
- Klyuev R.V., Gavrina O.A., Khetagurov V.N., Zaseev S.G., Umirov B.Z. Prediction of specific electric energy consumption at processing plant. *Mining Informational and Analytical Bulletin*, 2020, vol. 11, no. 1, pp. 135–145. In Rus.
- Klyuev R.V., Golik V.I., Bosikov I.I., Gavrina O.A. Analysis of electric power loss in the power supply system of the concentrating factory. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2021, vol. 332, no. 10, pp. 7–16. In Rus.
- Deqiang He, Xiaoliang Teng, Yanjun Chen, Bin Liu, Heliang Wang, Xianwang Li, Rui Ma. Energy saving in metro ventilation system based on multi-factor analysis and air characteristics of piston vent. *Applied Energy*, 2022, vol. 307, pp. 118295.
- Klyuev R.V., Bosikov I.I., Mayer A.V., Gavrina O.A. Comprehensive analysis of the effective technologies application to increase sustainable development of the natural-technical system. *Sustainable Development of Mountain Territories*, 2020, no. 2, pp. 283–290. In Rus.
- Da-sheng Lee, Yan-Tang Chen, Shih-Lung Chao. Universal workflow of artificial intelligence for energy saving. *Energy Reports*, 2022, vol. 8, pp. 1602–1633.
- Xiu'e Yang, Shuli Liu, Yuliang Zou, Wenjie Ji, Qunli Zhang, Abdullahi Ahmed, Xiaojing Han, Yongliang Shen, Shaoliang Zhang. Energy-saving potential prediction models for large-scale building: A state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 2022, vol. 156, pp. 111992.
- Wang R.Q., Jiang L., Wang Y.D., Roskilly A.P. Energy saving technologies and mass-thermal network optimization for decarbonized iron and steel industry: a review. *Journal of Cleaner Production*, 2020, vol. 274, pp. 122997.
- Tiancheng Shang, Peihong Liu, Junxiong Guo. How to allocate energy-saving benefit for guaranteed savings EPC projects? A case of China. *Energy*, 2020, vol. 191, pp. 116499.

11. Yu Liu, Guoyuan Ma, Lianzheng Xue, Feng Zhou, Lei Wang. Energy-saving effect of integrated cooling unit with rotary booster and compressor for data center. *International Journal of Refrigeration*, 2020, vol. 119, pp. 366–375.
12. Xinyue Li, Shuqin Chen, Hongliang Li, Yunxiao Lou, Jiahe Li. Multi-dimensional analysis of air-conditioning energy use for energy-saving management in university teaching buildings. *Building and Environment*, 2020, vol. 185, pp. 107246.
13. Garcia-Martín E., Faviola Rodrigues C., Riley G., Grahn H. Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, 2019, vol. 134, pp. 75–88.
14. Ghodduzi H., Creamer G.G., Rafizadeh N. Machine learning in energy economics and finance: a review. *Energy Economics*, 2019, vol. 81, pp. 709–727.
15. Shintaro Ikeda, Tatsuo Nagai. A novel optimization method combining metaheuristics and machine learning for daily optimal operations in building energy and storage systems. *Applied Energy*, 2021, vol. 289, pp. 116716.
16. *Prognozirovanie i planirovanie potrebleniya elektricheskoy energii s pomoshchyu machine learning (experiment)* [Forecasting and planning of electricity consumption using machine learning (experiment)]. Available at: <https://habr.com/ru/post/577732/> (accessed 9 January 2022).
17. Amasyalia K., El-Gohary N. Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. *Renewable and Sustainable Energy Reviews*, 2021, vol. 142, pp. 110714.
18. Faramarz Bagherzadeh, Amirreza Shojaei Nouri, Mohamad-Javad Mehrani, Suresh Thennadil. Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach. *Process Safety and Environmental Protection*, 2021, vol. 154, pp. 458–466.
19. Flick D., Keck C., Herrmann Ch., Thiede S. Machine learning based analysis of factory energy load curves with focus on transition times for anomaly detection. *Procedia CIRP*, 2020, vol. 93, pp. 461–466.
20. *CatBoost Regressor*. Available at: https://catboost.ai/en/docs/concepts/python-reference_catboostregressor (accessed 11 January 2022).

Received: 17 June 2022.

Information about the authors

Anzhelika D. Morgoeva, postgraduate student, North Caucasian Institute of Mining and Metallurgy (State Technological University).

Irbek D. Morgoev, postgraduate student, North Caucasian Institute of Mining and Metallurgy (State Technological University).

Roman V. Klyuev, Dr. Sc., professor, Moscow Polytechnic University.

Oksana A. Gavrina, Cand. Sc., assistant professor, North Caucasian Institute of Mining and Metallurgy (State Technological University).