ISSN 2782-2001
Системы анализа
и обработки данных
том 103, № 3, 2026, с. 115–130

http://journals.nstu.ru/vestnik
Analysis and data
processing systems
Vol. 103, No. 3, 2026, pp. 115–130

ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ
И ТЕЛЕКОММУНИКАЦИИ

INFORMATION
TECHNOLOGIES
AND TELECOMMUNICATIONS

УДК 303.732

DOI: 10.17212/2782-2001-2026-3-115-130

Анализ влияния гиперпараметров ансамблевых моделей на процесс их обучения в задаче прогнозирования электропотребления предприятия газовой промышленности*

А.И. СТЕПАНОВА[a], А.И. ХАЛЬЯСМАА[b], П.В. МАТРЕНИН[c]

РФ, 620062, г. Екатеринбург, ул. Мира, 19, Уральский федеральный университет имени первого Президента России Б.Н. Ельцина

[a] *a.i.stepanova@urfu.ru* [b]*a.i.khaliasmaa@urfu.ru* [c]*p.v.matrenin@urfu.ru*

В настоящей статье представлен анализ влияния гиперпараметров ансамблевых моделей на точность, стабильность и вычислительную эффективность в задаче краткосрочного прогнозирования потребления электрической энергии. Объектом исследования являлась реальная компрессорная станция, находящаяся на севере Центральной России. Эксперименты проводились для моделей, построенных на основании быстрого бустинга, экстремального градиентного бустинга, адаптивного бустинга и на основании случайного леса. В качестве гиперпараметров были выбраны эффективная глубина деревьев и количество базовых моделей. Общее количество комбинаций эффективной глубины деревьев и количества базовых деревьев составило 72 для каждой модели. Вклад работы заключается в проведении эксперимента по полному перебору комбинацией гиперпараметров на реальных данных предприятия газовой промышленности, оценки склонности моделей к переобучению и недообучению сравнением средней относительной ошибки по модулю на обучающей и тестовой выборках. Также были выявлены оптимальные конфигурации гиперпараметров для каждой модели с учетом компромисса между точностью, скоростью и стабильностью. По результатам экспериментов лучшей стабильностью обладают модели, построенные на основе случайного леса и адаптивного бустинга, но из-за высоких показателей времени обучения и низких показателей точности эти модели не могут быть применены для краткосрочного прогнозирования потребления предприятием газовой промышленности. Лучшими характеристиками точности и предсказуемости ошибки обладают модели, построенные на основе экстремального градиентного бустинга; хорошими показателями точности и самым низким временем обучения обладают модели, построенные на основе быстрого бустинга.

* *Статья получена 11 февраля 2026 г.*

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ (тема № FEUZ-2025–0005).

Ключевые слова: ансамблевые методы, дерево решений, эффективная глубина дерева, количество базовых моделей, прогнозирование электропотребления, предприятие газовой промышленности, склонность к недообучению, склонность к переобучению, точность обучения, скорость обучения, стабильность обучения

ВВЕДЕНИЕ

Повышение энергетической эффективности и управление ресурсами являются ключевыми приоритетами предприятий газовой промышленности. Одним из перспективных мероприятий для анализа показателей энергетической эффективности является создание и внедрение системы краткосрочного прогнозирования потребления электрической энергии (1–3 суток). Внедрение такой системы позволяет:

- выровнять график потребления электрической энергии [1, 2];
- оптимизировать режимы работы объектов потребления, генерации и накопления электрической энергии [3, 4];
- минимизировать затраты на использование потребляемой мощности за счет регулирования тарифов на оптовом рынке электрической энергии и мощности путем планирования нагрузки [1, 5].

Внедрение усложнено из-за необходимости учета сложной структуры энергетического комплекса предприятий и сильной зависимости потребления электрической энергии от технологических процессов, климатических условий и режимов эксплуатации оборудования [5].

Для краткосрочного прогнозирования электропотребления применяются различные методы, которые можно разделить на детерминированные (статистические) методы и методы машинного обучения. К первым относятся сезонные модели [6] и методы на основе авторегрессии [7–9].

Методы машинного обучения способны учитывать большое количество факторов, в том числе метеорологических [10, 11] и производственных [12], а также зависимости между ними [13]. Преимущества применения методов машинного обучения по сравнению с детерминированными показаны в работах [14, 15]. Как правило, наилучшие результаты достигаются с помощью нейросетевых моделей [16–18], включая рекуррентные [19] и глубокие нейронные сети [17, 20, 21], ансамбли деревьев решений [22].

Однако, несмотря на широкое применение этих методов, точность и обобщающая способность методов зависят от выбора гиперпараметров. Большинство существующих исследований используют эвристический метод выбора параметров, в то время как работы [23–25] показывают, что даже небольшие изменения в гиперпараметрах могут приводить к колебаниям точности и повышению времени обучения в условиях ограниченности данных и высокого уровня шума [26].

В настоящее время отсутствуют комплексные исследования, посвященные эмпирическому анализу чувствительности ансамблевых моделей к гиперпараметрам для предприятий газовой промышленности.

Настоящая работа является продолжением исследования прогнозирования потребления электрической энергии предприятия газовой промышленности [27]. Целью исследования является анализ влияния ключевых гиперпараметров на точность, стабильность и вычислительную эффективность

моделей, построенных на адаптивном бустинге, экстремальном градиентном бустинге, быстром бустинге, случайном лесе (бустинг от англ. *boosting* – метод формирования ансамблей моделей машинного обучения, при котором несколько относительно простых моделей обучаются последовательно так, чтобы каждая следующая модель уменьшала ошибки предыдущих).

Вклад работы заключается в следующем.

1. Оценена склонность каждой модели к переобучению и недообучению сравнением метрик на обучающей и тестовой выборках.

2. Выявлены оптимальные конфигурации гиперпараметров для каждой модели с учетом компромисса между точностью, скоростью и стабильностью.

3. Сформулированы практические рекомендации по выбору модели и ее настройке для предприятий газовой промышленности.

В разделе 1 представлено описание используемых материалов и методов. В разделе 2 показаны результаты исследования влияния гиперпараметров на результаты обучения моделей. В заключении представлено описание будущих направлений исследования.

1. ИСПОЛЬЗУЕМЫЙ МЕТОД КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ

В работе используются архивные данные электропотребления компрессорной станции магистрального газопровода и метеорологические параметры, применяемые в качестве входных признаков моделей машинного обучения для краткосрочного прогнозирования потребления электрической энергии.

Объектом исследования в статье является компрессорная станция магистрального газопровода, которую можно определить как комплекс оборудования для повышения давления газа и его охлаждения при транспортировке. Рассматриваемая в статье станция расположена на севере Центральной России и состоит из трех компрессорных цехов.

Исходный набор данных включал трехлетние значения (791 день):

- почасового потребления электроэнергии компрессорной станцией для всего предприятия, полученные с автоматизированной системы коммерческого учета;
- суточного потребления электроэнергии аппаратами воздушного охлаждения (АВО) трех установок охлаждения газа;
- суточного потребления электроэнергии газоперекачивающими агрегатами (ГПА) трех компрессорных цехов.

Также в экспериментах рассматривались метеорологические данные, которые были взяты из архива сайта rp5.ru. Вместо прогнозных данных использовался архив фактических метеоданных из-за отсутствия открытого доступа к архивам прогнозных значений. Такое упрощение может быть оправдано тем, что метеорологические параметры являются дополнительными факторами, а точность прогноза погоды сопоставима с фактическими параметрами погоды на краткосрочном горизонте.

Описание предобработки данных представлено в статье [27]. Выбросы в наборе данных отсутствовали.

После корреляционного анализа, проведенного в статье [27], прогнозируемое потребление электрической энергии описывается как

$$y_i^* = f(y_{i-48}, y_{i-60}, y_{i-72}, y_{i-84}, y_{i-96}, X),$$

где y_i^* – прогнозируемое потребление электрической энергии в час i; f – прогнозная модель; X – другие значения (метеорологические данные, данные производственного календаря, суточное потребление ГПА, АВО).

Следует отметить, что наличие мультиколлинеарности входных признаков потенциально может влиять на устойчивость моделей. Однако для ансамблевых методов на основе деревьев решений данный эффект частично нивелируется за счет случайного выбора признаков и структуры разбиений.

Весь набор данных был разделен на обучающие и тестовые наборы данных в соотношении 90:10. Для повышения достоверности полученных результатов дополнительно использовалась процедура кросс-валидации на обучающей выборке (с использованием $k = 5$ подвыборок), в рамках которой модель обучалась на $k - 1$ подвыборках и тестировалась на оставшейся выборке. Итоговые значения метрик определялись как усредненные по всем разбиениям.

Для решения задачи краткосрочного прогнозирования потребления электрической энергии в работе выбраны ансамблевые методы машинного обучения: случайный лес (СЛ), адаптивный бустинг (АБ) [28], экстремальный градиентный бустинг (ГБ) [29], быстрый бустинг (ББ) [30].

Для обеспечения сопоставимости и контроля сложности моделей в задаче оценки влияния параметров моделей на точность и скорость ее работы в настоящей работе были выбраны эффективная глубина дерева и количество базовых деревьев.

Эффективная глубина дерева определяет максимальную сложность базового решающего дерева. Для адаптивного бустинга, экстремального градиентного бустинга и случайных деревьев использовался гиперпараметр max_depth. Так как в быстром бустинге рост дерева контролируется числом листьев, для него использовался параметр num_leaves.

Количество базовых моделей определяет число деревьев в ансамбле и выражается n_estimators. Этот параметр является общим для всех рассматриваемых моделей.

Выбор параметров обусловлен влиянием на компромисс смещения и дисперсии (bias-variance tradeoff), а также влиянием на вычислительную сложность модели.

Установлены диапазоны значений гиперпараметров:

- max_depth (num_leaves): 3, 5, 10, 20, 30, 50, 70, 100;
- n_estimators: от 10, 30, 50, 70, 90, 100, 150, 200, 300.

При значении max_depth (num_leaves) до 10 анализируются слабые модели. При значении глубины больше 50 анализируются сложные модели, что позволяет выявить режимы недообучения и переобучения.

Минимальные значения n_estimators позволяют покрыть минимальные конфигурации для оценки нижней границы точности и насыщенные ансамбли для оценки предела улучшения модели.

Общее количество комбинаций – 72 для каждой модели.

Выбор дискретных значений гиперпараметров с относительно крупным шагом обусловлен необходимостью ограничения вычислительной сложности полного перебора при сохранении репрезентативности исследуемого пространства параметров.

Для сравнения результатов и оценки склонности каждой модели к переобучению и недообучению были рассчитаны время обучения, средняя абсолютная ошибка (MAE), средняя относительная ошибка по модулю (MAPE), коэффициент детерминации (R2).

Предобработка данных, построение и тестирование моделей машинного обучения было выполнено на языке программирования Python 3 с применением библиотек с открытым исходным кодом: Scikit-Learn (методы адаптивного бустинга, случайного леса) [28], XGBoost (метод экстремального градиентного бустинга) [29], LightGBM (метод быстрого бустинга) [30].

Эксперименты проводились на вычислительной системе со следующими характеристиками: центральный процессор Intel Core i5 G7, 4 ядра 2,4 ГГц, 16 ГБ оперативной памяти.

2. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

На рис. 1–4 представлены поверхности, которые показывают зависимость времени обучения модели от эффективной глубины и количества базовых моделей для экстремального градиентного бустинга, адаптивного бустинга, быстрого бустинга и случайного леса.

Поверхности строятся для наглядного анализа вычислительной эффективности, формы поверхности.

Наибольшей скоростью обучения обладают модели, построенные на основе быстрого бустинга. Это обосновано следующими особенностями метода:

- лист-ориентированный рост дерева;
- использование гистограммной бинаризации признаков;
- фильтрация малозначимых признаков.

Неровности формы поверхности быстрого бустинга обоснованы оптимизацией, зависящей от структуры данных, и динамическим выбором признаков.

Волнистость поверхности быстрого бустинга выражена слабо, что говорит о высокой эффективности оптимизации.

Время обучения моделей, построенных на основе адаптивного бустинга, линейно зависит от количества базовых моделей на всей поверхности. Зависимость от эффективной глубины дерева имеет линейный характер до max_depth = 20, после этого значения время обучения не изменяются при возрастании глубины дерева. Адаптивный бустинг обучается медленнее быстрого бустинга, так как деревья обучаются последовательно, пересчитывая веса объектов на каждом шаге без использования оптимизации.

Форма поверхности адаптивного бустинга отражает детерминированную сложность (сложность модели полностью определяется параметрами).

Время обучения моделей, построенных на основе случайного леса, линейно зависит от количества базовых моделей.

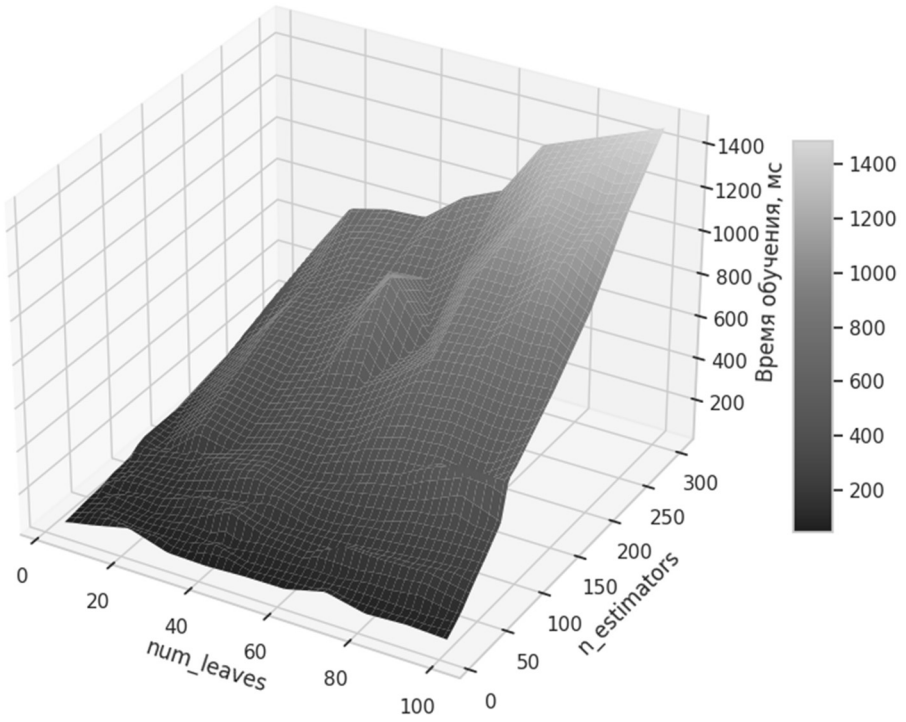


Рис. 1. Зависимость времени обучения от гиперпараметров для ББ

Fig. 1. Training time as a function of hyperparameters for LightGBM

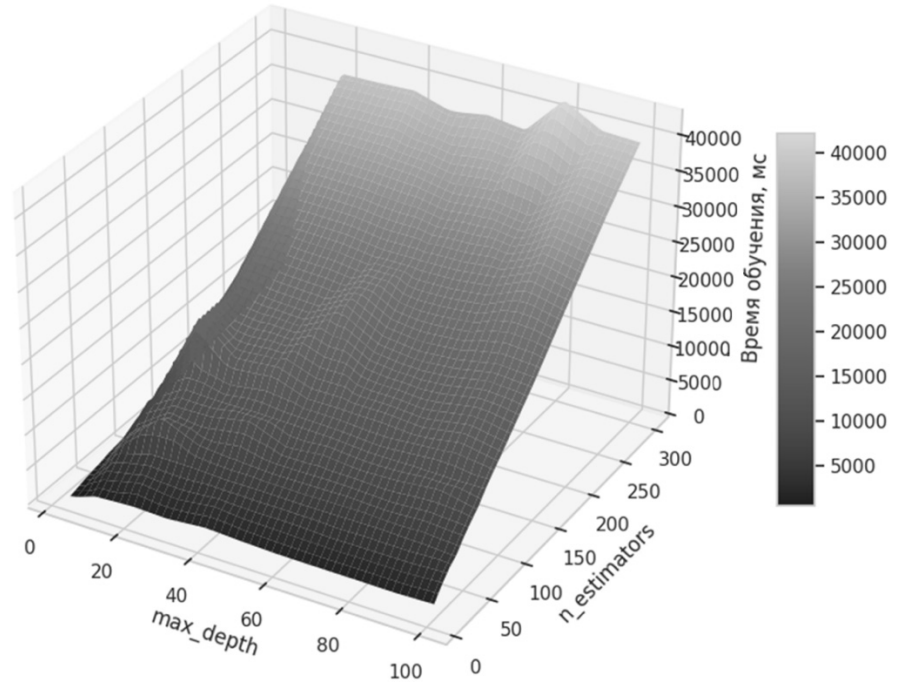


Рис. 2. Зависимость времени обучения от гиперпараметров для АБ

Fig. 2. Training time as a function of hyperparameters for AdaBoost

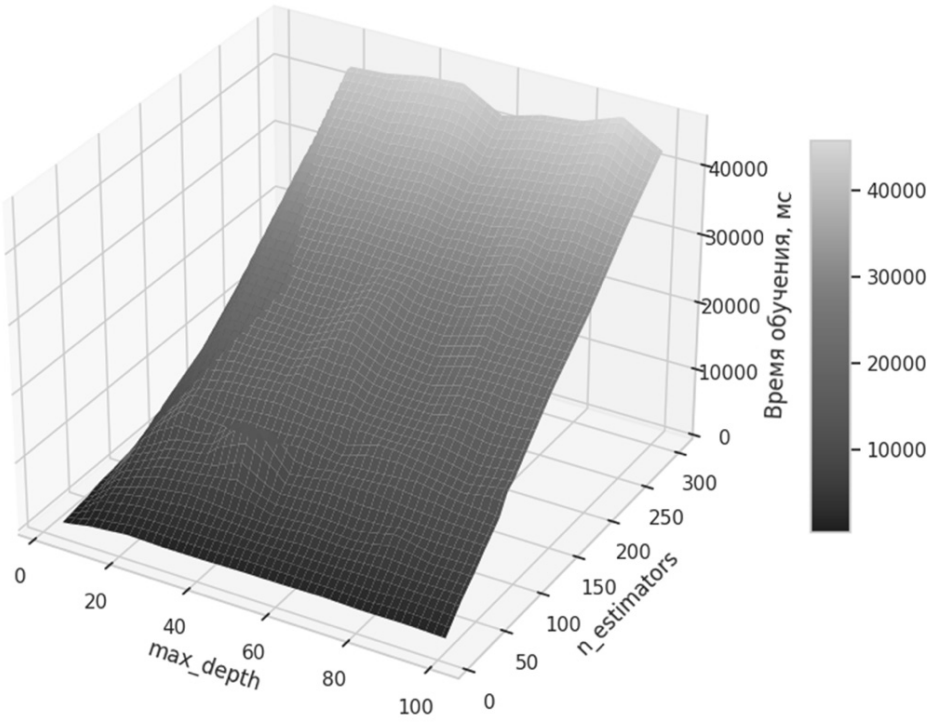


Рис. 3. Зависимость времени обучения от гиперпараметров для СЛ

Fig. 3. Training time as a function of hyperparameters for Random Forest

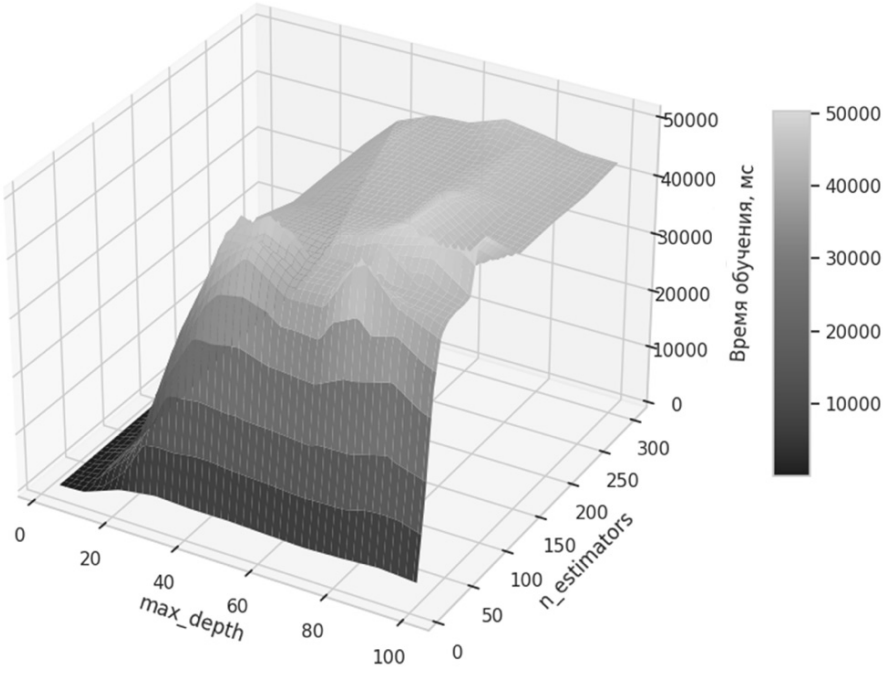


Рис. 4. Зависимость времени обучения от гиперпараметров для ГБ

Fig. 4. Training time as a function of hyperparameters for XGBoost

Зависимость от эффективной глубины дерева имеет линейный характер до max_depth = 20, после этого значения времени обучения не изменяются. Большее время обучения по сравнению с быстрым бустингом обосновано тем, что каждое дерево строится на полной выборке без оптимизации.

Гладкая поверхность случайного леса показывает стабильность работы модели при изменении параметров.

Время обучения моделей, построенных на основе экстремального градиентного бустинга, линейно зависит от эффективной глубины дерева до max_depth = 40, после этого не изменяется. При этом значения времени обучения для эффективной глубины больше max_depth = 40 совпадают. Время обучения модели экстремального градиентного бустинга определяется сочетанием глубины деревьев и количества базовых моделей. При значениях параметра max_depth < 30 увеличение количества базовых моделей практически не влияет на время обучения. При большей глубине деревьев (max_depth ≥ 30) и количестве базовых моделей до 100 время обучения возрастает линейно с увеличением их числа. При дальнейшем увеличении количества базовых моделей существенного изменения времени обучения не наблюдается. Время обучения экстремального градиентного бустинга имеет наибольшие значения, так как он использует точный поиск разбиений и вычисляет вторые производные функции потерь. Это обеспечивает более высокую точность, но сказывается на вычислительной сложности.

Неровность поверхности обоснована оптимизацией, зависящей от структуры данных, и кэшированием градиентов.

Сильная волнистость поверхности экстремального градиентного бустинга отображает высокую чувствительность к структуре данных при max_depth >30, что приводит к уменьшению предсказуемости времени обучения. Высокая чувствительность к гиперпараметрам может увеличить операционные риски при автоматическом развертывании.

Гладкость поверхностей адаптивного бустинга и случайного леса обеспечивает предсказуемость, но ценой низкой скорости. На рис. 5 представлен график зависимости MAPE от количества базовых моделей для анализа склонности моделей к переобучению. Сплошные линии соответствуют значению MAPE на тестовой выборке, пунктирные – на обучающей. Для повышения читаемости рисунка представлены зависимости только для четырех значений эффективной глубины дерева.

При анализе графика зависимости MAPE от количества базовых моделей для адаптивного бустинга можно заключить, что для всех значений max_depth наблюдается переобучение. С ростом количества базовых моделей ошибка тестовой выборки растет, что указывает на неустойчивость модели при увеличении сложности ансамбля.

График зависимости MAPE от количества базовых моделей для экстремального градиентного бустинга показывает наилучшую устойчивость. Так как разница MAPE между тестовой и обучающей выборками минимальная, можно заключить, что у экстремального градиентного бустинга низкая склонность к переобучению. Также видно, что MAPE незначительно зависит от эффективной глубины выборки.

Для графика зависимости MAPE от количества базовых моделей быстрого бустинга при значениях количества базовых моделей до 50 наблюдается резкое

падение MAPE, что говорит о быстрой адаптации модели. При количестве базовых моделей больше 100 MAPE выходит на плато, дальнейшее увеличение числа деревьев не улучшает точность. Зависимость от max_depth выражена слабо.

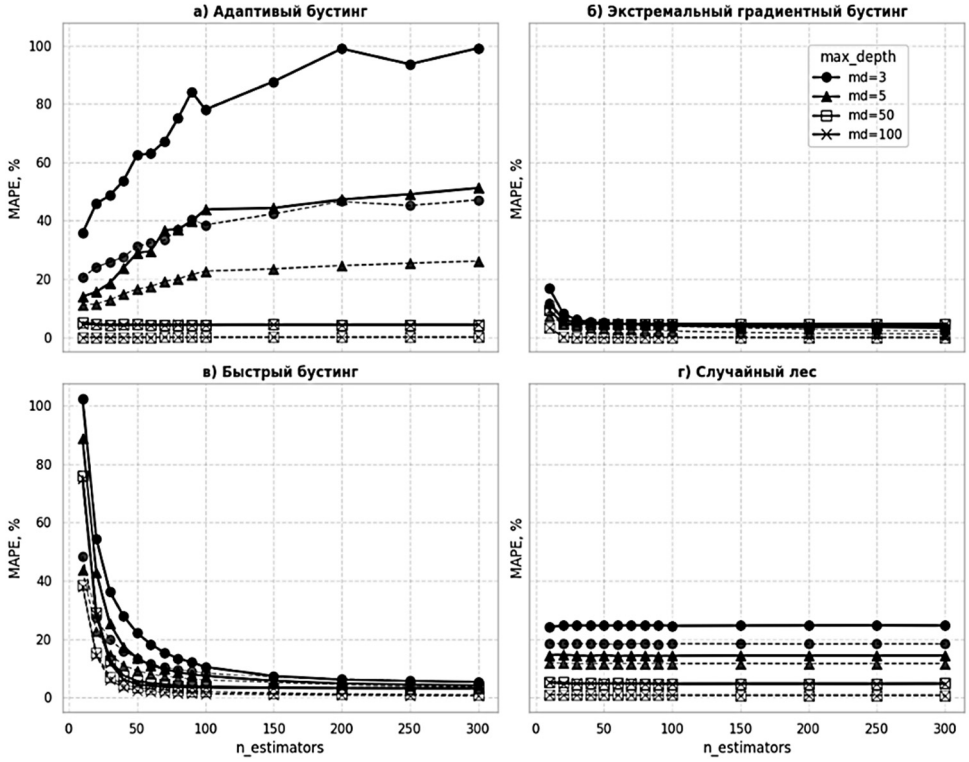


Рис. 5. Зависимость MAPE от количества базовых моделей для адаптивного бустинга (*а*), экстремального градиентного бустинга (*б*), быстрого бустинга (*в*), случайного леса (*г*)

Fig. 5. MAPE as a function of the total number of trees for AdaBoost (*a*), XGBoost (*b*), LightGBM (*c*), Random Forest (*d*)

Для графика зависимости MAPE от количества базовых моделей случайного леса MAPE находится в диапазоне 5…25 % в зависимости от эффективной глубины дерева. При эффективной глубине дерева 3 модель недообучена. Параметры точности моделей, построенных с использованием моделей случайного леса, стабилизируются при max_depth = 20. При значениях эффективной глубины дерева ошибка снижается и стабилизируется. Это значение является точкой оптимального компромисса между точностью и скоростью.

Сравнение лучших результатов обучения по точности и скорости для всех моделей представлено в таблице. Выбор оптимальной модели осуществлялся на основе многокритериального подхода, учитывающего одновременно точность, устойчивость (разница между обучающей и тестовой ошибкой) и вычислительную эффективность (время обучения). В рамках исследования применялся принцип Парето-оптимальности, при котором предпочтение отдавалось моделям, обеспечивающим наилучший компромисс между указанными критериями.

Модель, построенная на основе экстремального градиентного бустинга, показывает лучший результат: MAPE на тестовой выборке составляет 3,24 %, R2 составляет 0,98. Несмотря на то что использование такой модели нецелесообразно при частом повторном обучении из-за высокого и непредсказуемого времени обучения, так как в задаче краткосрочного прогнозирования потребления электрической энергии предприятия газовой промышленности нет необходимости часто обучать модель на новых данных, модель на основе экстремального градиентного бустинга может быть выбрана для применения на объекте.

Сравнение лучших результатов обучения

Comparison of the best training results

Модель	max_ depth	n_ estima- tors	MAE обуч, кВт	MAE тест, кВт	MAPE тест, %	R2	Время обучения, мс
ББ	5	300	23,11	13,40	4,24	0,96	324
ГБ	3	300	12,73	9,20	3,24	0,98	359
АД	10	100	13,79	21,92	7,75	0,91	9909
СЛ	30	100	5,73	16,02	4,78	0,93	15665

Также хорошими показателями обладает модель, построенная с использованием быстрого бустинга: MAPE на тестовой составляет 4,24 %, R2 составляет 0,96. При внедрении модели и выборе между моделями на основе экстремального градиентного бустинга и быстрого бустинга необходимо учитывать, что модели на основе быстрого бустинга демонстрируют высокую стабильность при правильных параметрах max_depth, что делает их оптимальными для промышленного внедрения при частом переобучении модели.

Случайный лес является наиболее предсказуемым по производительности, однако уступает в точности моделям на основе быстрого бустинга и экстремального градиентного бустинга. Лучшие показатели: MAPE на тестовой выборке составляет 4,78 % при времени обучения 15 665 мс.

Адаптивный бустинг склонен к переобучению. Для лучшего показателя время обучения модели меньше, чем у случайного леса, но склонность к переобучению и более выраженная по сравнению со случайным лесом нестабильность делает бустинг непрактичным для применения для рассматриваемой задачи.

С практической точки зрения снижение ошибки прогнозирования на уровне 1–2 процентных пункта MAPE является значимым для энергетических объектов, поскольку позволяет более точно планировать график потребления, снижать штрафы за отклонение от заявленных режимов и оптимизировать

загрузку оборудования. Для компрессорных станций это выражается в снижении эксплуатационных затрат и повышении энергоэффективности за счет более точного управления режимами работы.

ЗАКЛЮЧЕНИЕ

В настоящей статье для задачи краткосрочного прогнозирования потребления электрической энергии на основе данных потребления электрической энергии компрессорной станцией был проведен эксперимент по полному перебору комбинаций гиперпараметров для выбора модели с целью промышленного внедрения.

На основе анализа форм поверхностей времени обучения наиболее стабильными являются модели случайного леса и адаптивного бустинга, однако они демонстрируют худшие показатели сходимости MAPE по сравнению с методами экстремального градиентного бустинга и быстрого бустинга. Модели, построенные на основе быстрого бустинга, имеют высокие способности к адаптации и высокую скорость обучения. Предлагаемой к внедрению моделью является модель, построенная на экстремальном градиентном бустинге, так как она обеспечивает минимальную ошибку (3,24 %). Однако при необходимости частого переобучения или изменения постановки задачи для прогнозирования потребления в режиме реального времени необходимо рассмотреть применение быстрого бустинга.

При интерпретации полученных результатов необходимо учитывать ряд ограничений: качество прогнозирования существенно зависит от полноты и достоверности исходных данных, включая корректность измерений и отсутствие пропусков. Результаты получены на данных одного промышленного объекта.

В ходе дальнейших исследований планируется провести дообучение моделей, построенных на основе экстремального градиентного бустинга и быстрого бустинга, данными других объектов газовой промышленности с последующим анализом поведения моделей.

СПИСОК ЛИТЕРАТУРЫ

1. *Lee E., Kim J., Jang D.* Load profile segmentation for effective residential demand response program: Method and evidence from Korean pilot study // Energies. – 2020. – Vol. 13 (6). – P. 1348. – DOI: 10.3390/en13061348.

2. A short-term forecasting approach for regional electricity power consumption by considering its co-movement with economic indices / K. Li, Z. Yang, D. Li, Y.Y. Xing, W. Nai // IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC). – IEEE, 2020. – P. 551–555. – DOI: 10.1109/ITOEC49072.2020.9141928.

3. Renewable on-site power generation for manufacturing companies – Technologies, modeling, and dimensioning / J. Schulz, D. Leinmüller, A. Misik, M.F. Zaeh // Sustainability. – 2021. – Vol. 13 (7). – P. 3898. – DOI: 10.3390/su13073898.

4. Optimal maintenance policy for equipment submitted to multi-period leasing as a circular business model / A.B. Mabrouk, A. Chelbi, M.S. Aguir, S. Dellagi // Sustainability. – 2024. – Vol. 16 (12). – P. 5238. – DOI: 10.3390/su16125238.

5. *Stepanova A.I., Matrenin P.V.* New load forecasting ensemble model based on LightGBM for gas industry enterprises // Proceedings of the 2024 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT). – IEEE, 2024. – P. 1–4. – DOI: 10.1109/USBEREIT61901.2024.10583981.

6. Power load forecasting research based on neural network and Holt-winters method / J. Fan, X. Liu, Z. Li, X. Wang, S. Cao, J. Lei // IOP Conference Series: Earth and Environmental Science. – 2021. – Vol. 692. – P. 022120. – DOI: 10.1088/1755-1315/692/2/022120.

7. Short-term forecast of electricity load for LLC “Omsk Energy Retail Company” using neural network / V. Potapov, R. Khamitov, V. Makarov, A. Gritsay, I. Chervenchuk, D. Tyunkov // Proceeding of 2018 Dynamics of Systems, Mechanisms and Machines (Dynamics). – IEEE, 2018. – P. 1–5. – DOI: 10.1109/Dynamics.2018.8601430.

8. *Chodakowska E., Nazarko J., Nazarko L.* ARIMA models in electrical load forecasting and their robustness to noise // Energies. – 2021. – Vol. 14 (23). – P. 7952. – DOI: 10.3390/en14237952.

9. Blind Kalman filtering for short-term load forecasting / S. Sharma, A. Majumdar, V. Elvira, E. Chouzenoux // IEEE Transactions on Power Systems. – 2020. – Vol. 35. – P. 4916–4919. – DOI: 10.1109/TPWRS.2020.3018623.

10. *Caro E., Juan J., Nouhitehrani S.* Optimal selection of weather stations for electric load forecasting // IEEE Access. – 2023. – Vol. 11. – P. 42981–42990. – DOI: 10.1109/ACCESS.2023.3270933.

11. Regression model-based short-term load forecasting for university campus load / M. Madhukumar, A. Sebastian, X. Liang, M. Jamil, M.N.S.K. Shabbir // IEEE Access. – 2022. – Vol. 10. – P. 8891–8905. – DOI: 10.1109/ACCESS.2022.3144206.

12. *Sergeev N., Matrenin P.* Improving accuracy of machine learning based short-term load forecasting models with correlation analysis and feature engineering // Proceeding of 2023 IEEE 24th International Conference of Young Professionals in Electron Devices and Materials (EDM). – IEEE, 2023. – P. 1000–1004. – DOI: 10.1109/EDM58354.2023.10225058.

13. Load forecasting techniques and their applications in smart grids / H. Habbak, M. Mahmoud, K. Metwally, M.M. Fouda, M.I. Ibrahem // Energies. – 2023. – Vol. 16 (3). – P. 1480. – DOI: 10.3390/en16031480.

14. *Ryu S., Noh J., Kim H*. Deep neural network based demand side short term load forecasting // Energies. – 2017. – Vol. 10 (1). – P. 3. – DOI: 10.3390/en10010003.

15. *Szczepaniuk H., Szczepaniuk E.K*. Applications of artificial intelligence algorithms in the energy sector // Energies. – 2023. – Vol. 16 (1). – P. 347. – DOI: 10.3390/en16010347.

16. A short-term load forecasting method based on GRU-CNN hybrid neural network model / L. Wu, C. Kong, X. Hao, W. Chen // Mathematical Problems in Engineering. – 2020. – P. 1428104. – DOI: 10.1155/2020/1428104.

17. Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting / Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, Z. Zhu // IEEE Access. – 2019. – Vol. 7. – P. 88058–88071. – DOI: 10.1109/ACCESS.2019.2926137.

18. Short-term residential load forecasting based on LSTM recurrent neural network / W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang // IEEE Transactions on Smart Grid. – 2019. – Vol. 10. – P. 841–851. – DOI: 10.1109/TSG.2017.2753802.

19. *Cai M., Pipattanasomporn M., Rahman S*. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques // Applied Energy. – 2019. – Vol. 236 (3). – P. 1078–1088. – DOI: 10.1016/j.apenergy.2018.12.042.

20. A deep learning method for short-term residential load forecasting in smart grid / Y. Hong, Y. Zhou, Q. Li, W. Xu, X. Zheng // IEEE Access. – 2020. – Vol. 8. – P. 55785–55797. – DOI: 10.1109/ACCESS.2020.2981817.

21. Load forecasting techniques for power system: Research challenges and survey / N. Ahmad, Y. Ghadi, M. Adnan, M. Ali // IEEE Access. – 2022. – Vol. 10. – P. 71054–71090. – DOI: 10.1109/ACCESS.2022.3187839.

22. Forecasting planned electricity consumption for the united power system using machine learning / R.V. Klyuev, A.D. Morgoeva, O.A. Gavrina, I.I. Bosikov, I.D. Morgoev // Journal of Mining Institute. – 2023. – Vol. 261. – P. 392–402.

23. *Mohammadagha M*. Hyperparameter optimization strategies for tree-based machine learning models prediction: A comparative study of AdaBoost, decision trees, and random forest // SSRN Electronic Journal. – 2025. – DOI: 10.2139/ssrn.5226457.

24. *Dasgupta S., Sen J*. A comparative study of hyperparameter tuning methods // Data Science in Theory and Practice. – 2024. – DOI: 10.48550/arXiv.2408.16425.

25. A unified framework for tuning hyperparameters in clustering problems / X. Fan, Y. Yue, P. Sarkar, Y.X.R. Wang // arXiv. – 2019. – DOI: 10.48550/arXiv.1910.08018.

26. *Hidayaturrohman Q.A., Hanada E.* A comparative analysis of hyper-parameter optimization methods for predicting heart failure outcomes // Applied Science. – 2025. – Vol. 15 (6). – P. 3393. – DOI: 10.3390/app15063393.

27. Application of SHAP and multi-agent approach for short-term forecast of power consumption of gas industry enterprises / A.I. Stepanova, A.I. Khalyasmaa, P.V. Matrenin, S.A. Eroshenko // Algorithms. – 2024. – Vol. 17 (10). – P. 447. – DOI: 10.3390/a17100447.

28. Scikit-Learn libraries. – URL: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning (accessed: 17.01.2026).

29. XGBoost. Introduction to Boosted Trees. – URL: https://xgboost.readthedocs.io/en/stable/tutorials/model.html (accessed: 17.01.2026).

30. LightGBM. – URL: https://lightgbm.readthedocs.io/en/stable/ (accessed: 17.01.2026).

Степанова Алина Игоревна, аспирант, младший научный сотрудник научной лаборатории цифровых двойников в электроэнергетике Уральского энергетического института, Уральский федеральный университет имени первого Президента России Б.Н. Ельцина. Основное направление научных исследований – методы искусственного интеллекта в электроэнергетике, объяснимый искусственный интеллект, анализ временных рядов. Имеет более 10 печатных работ. E-mail: a.i.stepanova@urfu.ru

Хальясмаа Александра Ильмаровна, кандидат технических наук, доцент, заведующий научной лабораторией цифровых двойников в электроэнергетике Уральского энергетического института, Уральский федеральный университет имени первого Президента России Б.Н. Ельцина. Основное направление научных исследований – техническая диагностика электрооборудования, методы искусственного интеллекта в электроэнергетике, робототехника. Имеет более 250 печатных работ. E-mail: a.i.khaliasmaa@urfu.ru

Матренин Павел Викторович, кандидат технических наук, доцент, ведущий научный сотрудник научной лаборатории цифровых двойников в электроэнергетике Уральского энергетического института, Уральский федеральный университет имени первого Президента России Б.Н. Ельцина. Основное направление научных исследований – методы искусственного интеллекта в электроэнергетике, объяснимый искусственный интеллект, анализ временных рядов. Имеет более 250 печатных работ. E-mail: p.v.matrenin@urfu.ru

Stepanova Alina I., PhD student, research assistant at the scientific laboratory of digital twins in the power industry, Ural Power Engineering Institute, Ural Federal University named after the first President of Russia, B.N. Yeltsin. The main field of her scientific research is methods of artificial intelligence in power industry, eXplainable artificial intelligence, and time series analysis. She is the author of more than 10 scientific publications. E-mail: a.i.stepanova@urfu.ru

Khalyasmaa Alexandra I., Ph.D, Head of scientific laboratory of digital twins in the power industry, Ural Power Engineering Institute, Ural Federal University named after the first President of Russia, B.N. Yeltsin. The main field of her scientific research is technical diagnostics of electrical equipment, methods of artificial intelligence in power industry, and robotics. She is the author of more than 250 scientific publications. E-mail: a.i.khaliasmaa@urfu.ru

Matrenin Pavel V., Ph.D, associate professor, leading researcher at the scientific laboratory of digital twins in the power industry, Ural Power Engineering Institute, Ural Federal University named after the first President of Russia, B.N. Yeltsin. The main field of his scientific research is methods of artificial intelligence in power industry, eXplainable artificial intelligence, and time series analysis. He is the author of more than 250 scientific publications. E-mail: p.v.matrenin@urfu.ru

DOI: 10.17212/2782-2001-2026-3-115-130

*Analysis of the impact of hyperparameters of ensemble models on their training process in the task of predicting electricity consumption for a gas industry enterprise**

A.I. STEPANOVA [a], A.I. KHALYASMAA [b], P.V. MATRENIN [c]

Ural Federal University, named after the first President of Russia, B.N. Yeltsin, 19, Mira Street, Ekaterinburg, 620062, Russian Federation

[a] *a.i.stepanova@urfu.ru* [b] *a.i.khaliasmaa@urfu.ru* [c] *p.v.matrenin@urfu.ru*

Abstract

This paper presents an analysis of the impact of hyperparameters on the accuracy, stability, and computational efficiency of ensemble models in short-term electrical energy consumption forecasting. The study was conducted on a real compressor station located in northern Central Russia. Experiments were performed using models based on LightGBM, XGBoost, AdaBoost, and Random Forest. The effective tree depth and the number of base learners were selected as hyperparameters, yielding 72 combinations of these parameters for each model. The contribution of this work lies in conducting a full hyperparameter grid search on real data from a gas industry enterprise assessing model tendencies toward overfitting and underfitting by comparing a mean absolute percentage error on training and test sets, and identifying optimal hyperparameter configurations for each model that balance accuracy, speed, and stability. Experimental results show that Random Forest and AdaBoost models demonstrate the highest stability; however, due to long training times and lower accuracy, these models are unsuitable for short-term forecasting in the gas industry applications. XGBoost-based models achieve the best accuracy and predictable error behavior, while LightGBM-based models provide competitive accuracy with the shortest training time.

Keywords: ensemble methods, decision tree, effective tree depth, number of base learners, electricity consumption forecasting, gas industry enterprise, underfitting tendency, overfitting tendency, prediction accuracy, training speed, training stability

REFERENCES

1. Lee E., Kim J., Jang D. load profile segmentation for effective residential demand response program: Method and evidence from Korean pilot study. *Energies*, 2020, vol. 13 (6), p. 1348. DOI: 10.3390/en13061348.

2. Li K., Yang Z., Li D., Xing Y.Y., Nai W. A short-term forecasting approach for regional electricity power consumption by considering its co-movement with economic indices. *IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*. IEEE, 2020, pp. 551–555. DOI: 10.1109/ITOEC49072.2020.9141928.

3. Schulz J., Leinmüller D., Misik A., Zaeh M.F. Renewable on-site power generation for manufacturing companies – Technologies, modeling, and dimensioning. *Sustainability*, 2021, vol. 13 (7), p. 3898. DOI: 10.3390/su13073898.

4. Mabrouk A.B., Chelbi A., Aguir M.S., Dellagi S. Optimal maintenance policy for equipment submitted to multi-period leasing as a circular business model. *Sustainability*, 2024, vol. 16 (12), p. 5238. DOI: 10.3390/su16125238.

5. Stepanova A.I., Matrenin P.V. New load forecasting ensemble model based on LightGBM for gas industry enterprises. *Proceedings of the 2024 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT)*. IEEE, 2024, pp. 1–4. DOI: 10.1109/USBEREIT61901.2024.10583981.

* *Received 11 February 2026.*

The work was carried out within the framework of a state assignment with financial support from the Ministry of Science and Higher Education of the Russian Federation (topic No. FEUZ-2025–0005).

6. Fan J., Liu X., Li Z., Wang X., Cao S., Lei J. Power load forecasting research based on neural network and Holt-winters method. *IOP Conference Series: Earth and Environmental Science*, 2021, vol. 692, p. 022120. DOI: 10.1088/1755-1315/692/2/022120.

7. Potapov V., Khamitov R., Makarov V., Gritsay A., Chervenchuk I., Tyunkov D. Short-term forecast of electricity load for LLC "Omsk Energy Retail Company" using neural network. *Proceeding of 2018 Dynamics of Systems, Mechanisms and Machines (Dynamics)*. IEEE, 2018, pp. 1–5. DOI: 10.1109/Dynamics.2018.8601430.

8. Chodakowska E., Nazarko J., Nazarko L. ARIMA models in electrical load forecasting and their robustness to noise. *Energies*, 2021, vol. 14 (23), p. 7952. DOI: 10.3390/en14237952.

9. Sharma S., Majumdar A., Elvira V., Chouzenoux E. Blind Kalman filtering for short-term load forecasting. *IEEE Transactions on Power Systems*, 2020, vol. 35, pp. 4916–4919. DOI: 10.1109/TPWRS.2020.3018623.

10. Caro E., Juan J., Nouhitehrani S. Optimal selection of weather stations for electric load forecasting. *IEEE Access*, 2023, no. 11, pp. 42981–42990. DOI: 10.1109/ACCESS.2023.3270933.

11. Madhukumar M., Sebastian A., Liang X., Jamil M., Shabbir M.N.S.K. Regression model-based short-term load forecasting for university campus load. *IEEE Access*, 2022, no. 10, pp. 8891–8905. DOI: 10.1109/ACCESS.2022.3144206.

12. Sergeev N., Matrenin P. Improving accuracy of machine learning based short-term load forecasting models with correlation analysis and feature engineering. *Proceeding of 2023 IEEE 24th International Conference of Young Professionals in Electron Devices and Materials (EDM)*, 2023, pp. 1000–1004. DOI: 10.1109/EDM58354.2023.10225058.

13. Habbak H., Mahmoud M., Metwally K., Fouda M.M., Ibrahem M.I. Load forecasting techniques and their applications in smart grids. *Energies*, 2023, vol. 16 (3), p. 1480. DOI: 10.3390/en16031480.

14. Ryu S., Noh J., Kim H. Deep neural network based demand side short term load forecasting. *Energies*, 2017, vol. 10 (1), p. 3. DOI: 10.3390/en10010003.

15. Szczepaniuk H., Szczepaniuk E.K. Applications of artificial intelligence algorithms in the energy sector. *Energies*, 2023, vol. 16 (1), p. 347. DOI: 10.3390/en16010347.

16. Wu L., Kong C., Hao X., Chen W. A short-term load forecasting method based on gru-cnn hybrid neural network model. *Mathematical Problems in Engineering*, 2020, p. 1428104. DOI: 10.1155/2020/1428104.

17. Deng Z., Wang B., Xu Y., Xu T., Liu C., Zhu Z. Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting. *IEEE Access*, 2019, vol. 7, pp. 88058–88071. DOI: 10.1109/ACCESS.2019.2926137.

18. Kong W., Dong Z., Jia Y., Hill D.H., Xu Y., Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, 2019, vol. 10, pp. 841–851. DOI: 10.1109/TSG.2017.2753802.

19. Cai M., Pipattanasomporn M., Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied Energy*, 2019, vol. 236 (3), pp. 1078–1088. DOI: 10.1016/j.apenergy.2018.12.042.

20. Hong Y., Zhou Y., Li Q., Xu Q., Zhou Y., Li Q. A deep learning method for short-term residential load forecasting in smart grid. *IEEE Access*, 2020, vol. 8, pp. 55785–55797. DOI: 10.1109/ACCESS.2020.2981817.

21. Ahmad N., Ghadi Y., Adnan M., Ali M. Load forecasting techniques for power system: Research challenges and survey. *IEEE Access*, 2022, vol. 10, pp. 71054–71090. DOI: 10.1109/ACCESS.2022.3187839.

22. Klyuev R.V., Morgoeva A.D., Gavrina O.A., Bosikov I.I., Morgoev I.D. Forecasting planned electricity consumption for the united power system using machine learning. *Journal of Mining Institute*, 2023, vol. 261, pp. 392–402.

23. Mohammadagha M. Hyperparameter optimization strategies for tree-based machine learning models prediction: A comparative study of AdaBoost, decision trees, and random forest. *SSRN Electronic Journal*, 2025. DOI: 10.2139/ssrn.5226457.

24. Dasgupta S., Sen J. A comparative study of hyperparameter tuning methods. *Data Science in Theory and Practice*, 2024. DOI: 10.48550/arXiv.2408.16425.

25. Xinjie F., Yuguang Y., Sarkar P., Wang Y.X.R. A unified framework for tuning hyperparameters in clustering problems. *arXiv*, 2019. DOI: 10.48550/arXiv.1910.08018.

26. Hidayaturrohman Q.A., Hanada E. A comparative analysis of hyper-parameter optimization methods for predicting heart failure outcomes. *Applied Science*, 2025, vol. 15 (6), p. 3393. DOI: 10.3390/app15063393.

27. Stepanova A.I., Khalyasmaa A.I., Matrenin P.V., Eroshenko S.A. Application of SHAP and multi-agent approach for short-term forecast of power consumption of gas industry enterprises. *Algorithms*, 2024, vol. 17 (10), p. 447. DOI: 10.3390/a17100447.

28. *Scikit-Learn libraries*. Available at: https://scikit-learn.org/stable/supervised_learning.html#supervised-learning (accessed 17.01.2026).

29. XGBoost. *Introduction to Boosted Trees*. Available at: https://xgboost.readthedocs.io/en/stable/tutorials/model.html (accessed 17.01.2026).

30. LightGBM. Available at: https://lightgbm.readthedocs.io/en/stable/ (accessed 17.01.2026).

Для цитирования:

Степанова А.И., Хальясмаа А.И., Матренин П.В. Анализ влияния гиперпараметров ансамблевых моделей на процесс их обучения в задаче прогнозирования электропотребления предприятия газовой промышленности // Системы анализа и обработки данных. – 2026. – № 3 (103). – С. 115–130. – DOI: 10.17212/2782-2001-2026-3-115-130.

For citation:

Stepanova A.I., Khalyasmaa A.I., Matrenin P.V. Analiz vliyaniya giperparametrov ansamblevykh modelei na protsess ikh obucheniya v zadache prognozirovaniya elektropotrebleniya predpriyatiya gazovoi promyshlennosti [Analysis of the impact of hyperparameters of ensemble models on their training process in the task of predicting electricity consumption for a gas industry enterprise]. *Sistemy analiza i obrabotki dannykh = Analysis and Data Processing Systems*, 2026, no. 3 (103), pp. 115–130. DOI: 10.17212/2782-2001-2026-3-115-130.