

## **СРАВНЕНИЕ МЕТОДОВ НА ОСНОВЕ ДЕРЕВЬЕВ РЕШЕНИЙ В ЗАДАЧЕ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ**

**А.Ю. Горшенин**

аспирант, e-mail: augomgtu@gmail.com

Омский государственный технический университет, Омск, Россия

**Аннотация.** Рассмотрены вопросы краткосрочного прогнозирования электропотребления с использованием методом машинного обучения на основе деревьев решений. Проведен спектральный анализ, выявлены основные периоды в данных. Выполнено сравнение эффективности методов машинного обучения при помощи расчета основных метрик регрессионных моделей прогнозирования.

**Ключевые слова:** прогнозирование электропотребления, машинное обучение, деревья решений, спектральный анализ.

### **Введение**

Прогнозирование электропотребления становится важным аспектом в России. Наблюдается значительная динамика в использовании электроэнергии в стране, которая является одним из ведущих производителей [1–4] и потребителей [5] энергии в мире. В начале 2023 г. ОАО «Системный оператор Единой энергетической системы» (ОАО «СО ЕЭС») опубликовал отчет о производстве и использовании электроэнергии [6].

Согласно представленным данным, общее потребление электроэнергии в Единой энергетической системе России в 2022 г. увеличилось на 1,36 % по сравнению с 2021 г., достигнув отметки в 1,06 трлн кВт\*ч. Сведения от СО ЕЭС также указывают на то, что производство электроэнергии в первой ценовой зоне энергорынка (включая европейскую часть России и Урал, составляющие примерно 80 % всего производства страны) по итогам 2022 г. практически не изменилось, уменьшившись лишь на 0,2 %, достигнув отметки примерно 850 млрд кВт\*ч. Тепловые электростанции являлись основной причиной снижения производства.

Во второй ценовой зоне (Сибирь) производство электроэнергии, наоборот, стабильно увеличивалось. За 2022 г. показатель вырос более чем на 1 %, превысив отметку в 219 млрд кВт\*ч. В регионе Сибири наблюдалось снижение загрузки гидроэлектростанций (более чем на 10 % по сравнению с предыдущим годом) в связи с рекордно низким уровнем воды, но производство тепловых электростанций выросло почти на треть, по сравнению с предыдущим годом. Крупные компании увеличивали производство, отвечая на увеличивающийся спрос в регионе.

Необходимость более точного краткосрочного прогнозирования электропотребления в России возрастает [7, 8]. Такие прогнозы необходимы для обеспечения стабильной и эффективной национальной энергосистемы [9], оптимизации производства энергии и эффективного распределения ресурсов. Методы машинного обучения получили распространение в этой области благодаря их способности обрабатывать большие и сложные наборы данных [10, 11], что позволяет выявлять сложные модели потребления и прогнозировать будущий спрос.

Целью исследования является сравнение эффективности методов Catboost [12], XGBoost [13], LightGBM (LGBM) [14] и Random Forest [15] в прогнозировании электропотребления.

## 1. Спектральный анализ ретроспективных данных об электропотреблении

Ретроспективные данные об электропотреблении на энергосбытовых предприятиях поступают из внешних источников. Это связано с применением специфических информационных систем [16], используемых для автоматизации операций на таких предприятиях. При построении краткосрочного прогноза первичные данные необходимо первоначально извлечь из источников и выполнить их предварительную обработку.

Кроме того, при проведении исследований важно учитывать как явные, так и скрытые зависимости в структуре этих данных. Для поиска таких закономерностей и регулярных колебаний наиболее подходящими моделями являются те, которые основаны на быстром преобразовании Фурье.

Для проводимого исследования использовались данные об электропотреблении за период с 2015 по 2023 гг. Фрагмент таких данных представлен в таблице 1. Для более наглядного анализа полученных результатов быстрого преобразования Фурье целесообразно представить их в виде периодограммы. Это позволит визуально выявить наличие постоянных составляющих или низкочастотных компонентов, таких как тренды [17].

Таблица 1. Фрагмент протокола измерений данных

Дата, время	Электропотребление, мВт*ч	Температура воздуха, °С
01.12.2021, 0:00	504,3	-8,3
01.12.2021, 1:00	483,7	-7,1
01.12.2021, 2:00	469,2	-6,0
...	...	...
30.06.2023, 22:00	310,9	8,2
30.06.2023, 23:00	286,8	7,7

При анализе исходных данных с использованием метода спектрального анализа при обнаружении трендовых компонент целесообразно исключить их из набора

данных. Это важно, так как наличие трендов может внести значительные ошибки в высокочастотной области периодограммы. В таких ситуациях обычно используется формальная модель, которая аппроксимирует тренд с использованием линейной комбинации различных степенных полиномов. В эту модель включается свободный член, и параметры тренда оптимизируются с использованием метода наименьших квадратов [18]. Затем полученные значения тренда вычитаются из исходных данных. В самом простом случае это означает исключение постоянной составляющей из данных (центрирование ряда). Среднее значение ряда можно найти с помощью следующей формулы:

$$m = \frac{1}{N} \sum_{k=0}^{N-1} x_k, \quad (1)$$

где  $N$  – количество элементов в ряде;  $x_k$  –  $k$ -ый элемент ряда.

Центрированный ряд из сходного получается следующим образом:

$$x_k^0 = x_k - m, \quad k = 0, 1, \dots, N - 1. \quad (2)$$

Отсутствие постоянной компоненты в центрированном временном ряде подтверждается с помощью графического представления данных.

Для вычисления периодограммы используется быстрое преобразование Фурье. Для использования быстрого преобразования исходный временной ряд нужно дополнить нулями таким образом, чтобы длина нового ряда была  $N_1 = 2^p N$ . С помощью быстрого преобразования Фурье для отдельно взятого ряда было получено:

$$X_j = FFT_j \{x_k^0\}_{k=0}^{N_1-1} = \sum_{k=0}^{N_1-1} x_k^0 e^{-i \frac{2\pi}{N_1} k j}, \quad j = 0, 1, \dots, N_1 - 1, \quad (3)$$

где  $i$  – комплексное число.

Тогда периодограммы ряда могут быть вычислены по формуле:

$$D_j = \frac{1}{N^2} [(Re X_j)^2 + (Im X_j)^2], \quad j = 0, 1, \dots, \frac{N_1}{2}. \quad (4)$$

Отсчёты периодограммы соответствуют частотам:

$$\nu_j = \Delta \nu_j, \quad j = 0, 1, \dots, \frac{N_1}{2}, \quad (5)$$

где  $\Delta \nu_j = \frac{1}{N_1 \Delta t}$ ,  $\Delta t$  – временной шаг ряда.

Для оценивания дисперсии необходимо использовать формулу:

$$\sigma_0^2 = \frac{1}{N-1} \sum_{k=0}^{N-1} (x_k^0)^2. \quad (6)$$

Используя дисперсию временного ряда, в дальнейшем можно рассчитать пороговое значение для обнаружения сигнала. По представленному алгоритму были получены график электропотребления (рис. 1) и фрагмент периодограммы с ограничением оси  $X$  до 100 единиц (рис. 2).

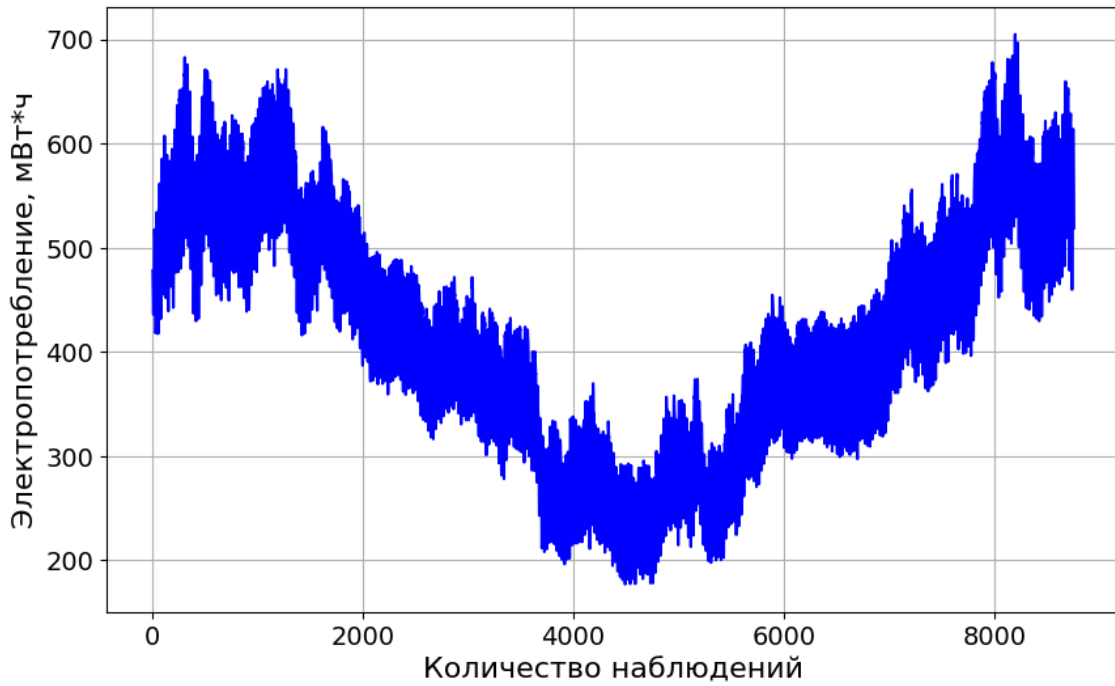


Рис. 1. График почасового электропотребления

Из графика установлена ярко выраженная цикличность процесса электропотребления.

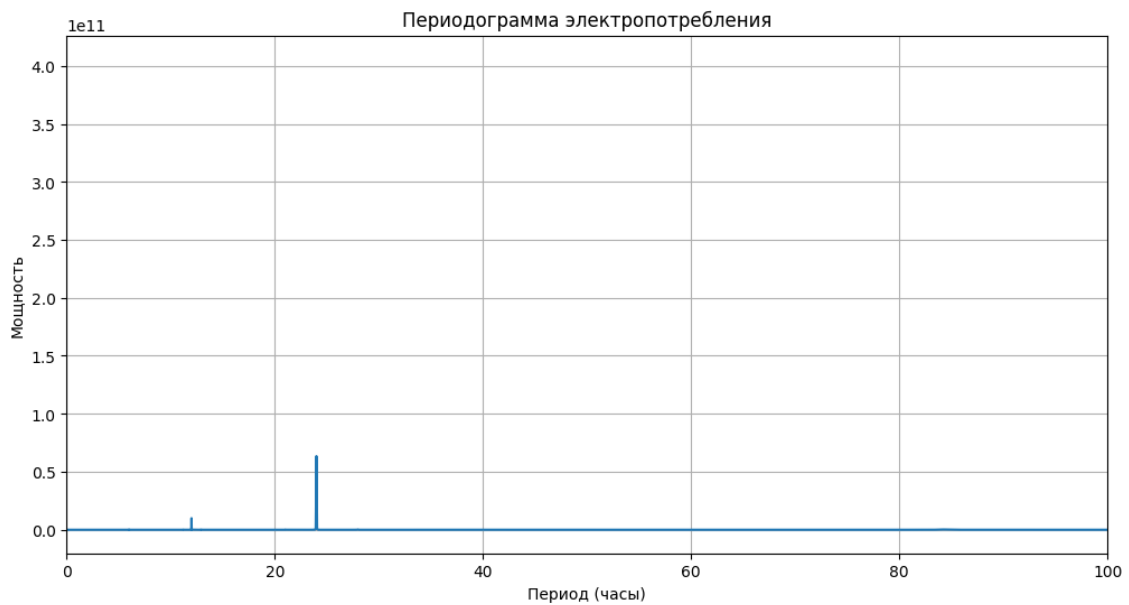


Рис. 2. Периодограмма почасового электропотребления

Далее была получена таблица всех пиков периодограммы, представленная в табл. 2.

Таблица 2. Пиковые значения периодограммы

№ п/п	Период, час	Периодограмма
1	8760,0000	4,058686e+11
2	4380,0000	6,535088e+08
3	2920,0000	1,305160e+09
4	2190,0000	8,039899e+09
...	...	...
4378	2,0009	1,174236e+03
4379	2,0004	3,017624e+05

В табл. 3 представлены наиболее мощные частотные характеристики. Проанализировав выделенные периоды (см. табл. 3.3), можно предположить, что сигналы под номерами 1 и 4 соответствуют году и трём месяцам соответственно, под номером 2 – суткам, сигнал под номером 3 соответствует половине месяца.

Таблица 3. Мощные частотные характеристики

№ п/п	Период, час	Периодограмма
1	8760,0000	4,058686e+11
2	24,0000	6,352815e+10
3	372,0000	9,964303e+09
4	2190,0000	8,039899e+09

Из приведённых выше данных видно, что были выявлены два важных цикла: суточный и годовой. Однако при более тщательном анализе мы можем обнаружить циклы продолжительностью, равной половине месяца и трём месяцам, и это указывает на наличие регулярных закономерностей в данных, которые следует учитывать при разработке краткосрочного прогноза для потребления электроэнергии.

## 2. Сравнение методов прогнозирования

Циклические данные демонстрируют регулярные повторяющиеся паттерны, или тенденции, во времени. Прогностическая пригодность циклических данных базируется на их повторяющихся образцах, которые могут быть идентифицированы и использованы для прогнозирования будущих событий. Для оценки эффективности прогнозирования электропотребления были использованы четыре основных метода машинного обучения: Catboost, XGBoost, LGBM и Random Forest. В исследовании использовались данные, подробно описанные в [7]. Графики рассеивания, представленные на рис. 3, отражают отношения между фактическими и прогнозируемыми значениями электропотребления при использовании каждого из методов.

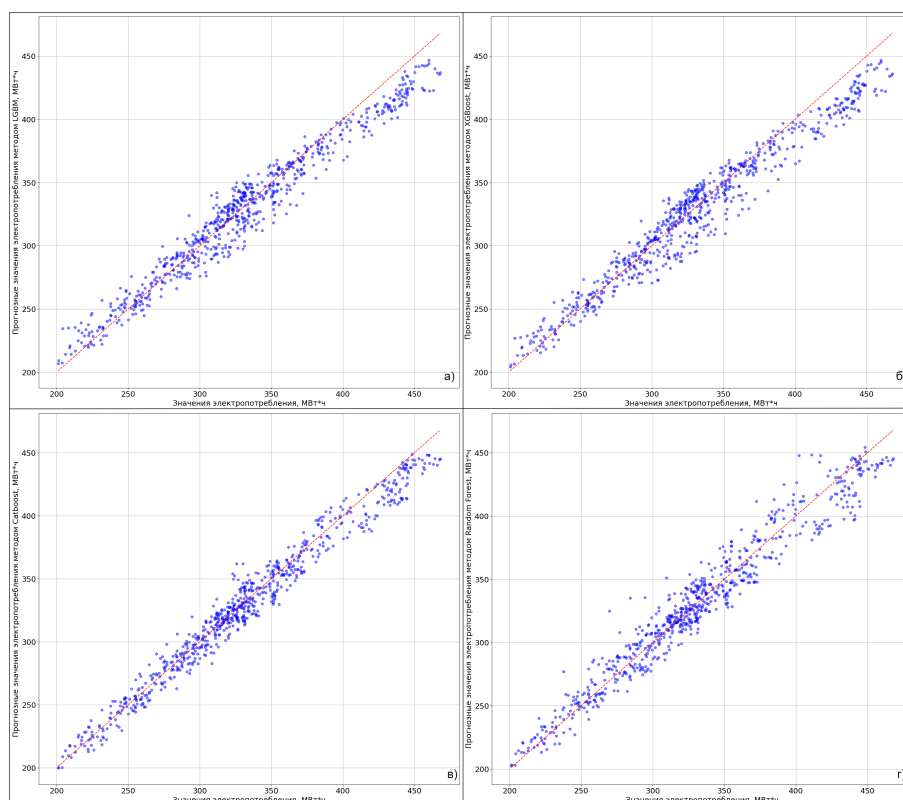


Рис. 3. Диаграмма рассеяния прогнозируемых и фактических значений моделей машинного обучения для прогнозирования активной мощности:  
(а) LGBM, (б) XGBoost, (в) Catboost, (г) Random Forest

Исходя из анализа графиков рассеивания, мы пришли к выводу, что метод Catboost демонстрирует наилучшую точность прогнозирования электропотребления по сравнению с XGBoost, LGBM и Random Forest. Его способность уменьшать разброс между фактическими и прогнозируемыми значениями электропотребления делает его наиболее надёжным инструментом для прогнозирования.

Рис. 4 иллюстрирует качество прогнозирования электропотребления за период с 01.06.2023 г. по 30.06.2023 г. На рисунке видно, что все методы имеют невысокий процент ошибки на месячном интервале, но лучше других себя показал метод Catboost (MAPE = 2,53 %).

На рис. 5 приведена гистограмма ошибок прогнозирования электропотребления для вариантов четырёх методов прогнозирования.

Рассматривался отрезок времени с 01.06.2023 г. по 30.06.2023 г., на котором видно, что наименьшие значения метрики MAPE имеет модель метода Catboost. Остальные методы показывают всплески ошибок в некоторые дни, что говорит о нестабильности этих методов и подтверждает обобщающую способность Catboost улавливать закономерности изменения целевой переменной.

Для оценивания качества работы и производительности представленных методов краткосрочного прогнозирования применялись метрики, характеризующие точность модели на фактических и прогнозных значениях [1].

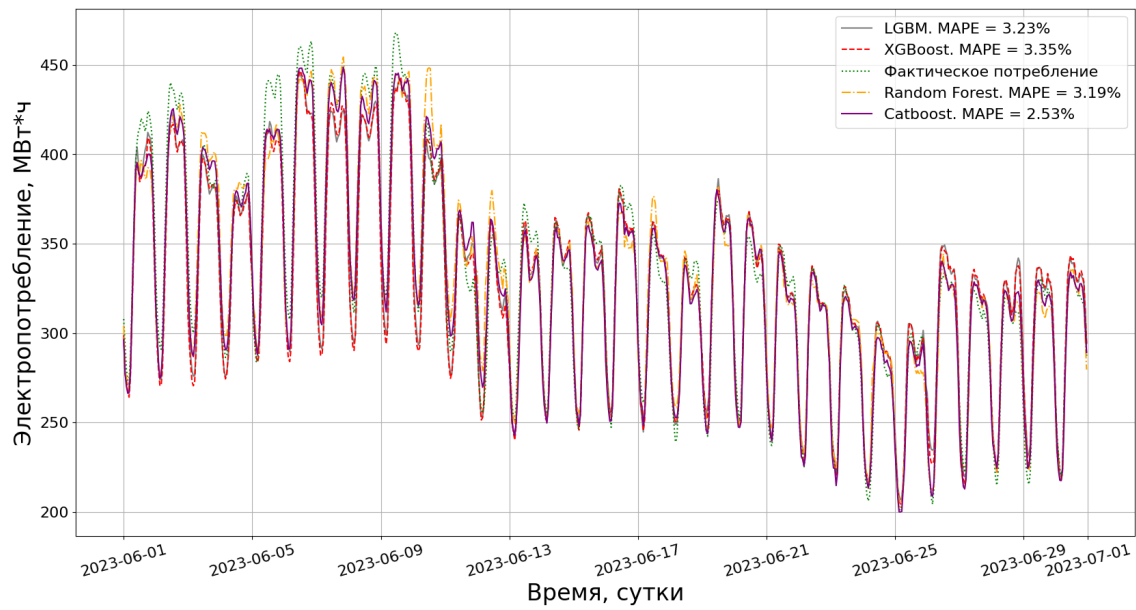


Рис. 4. Распределение электропотребления

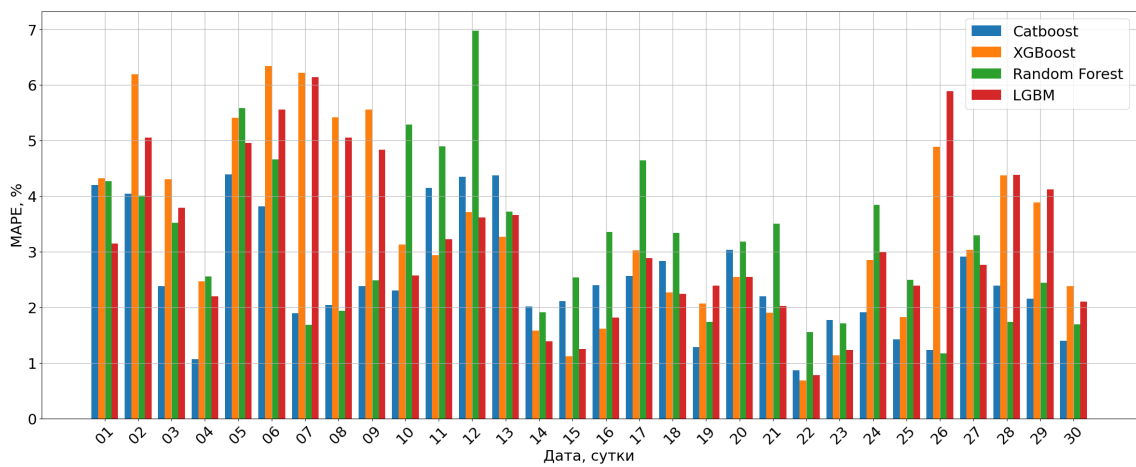


Рис. 5. Гистограмма ошибок (MAPE) прогнозирования электропотребления

Таблица 4. Метрики качества модели прогнозирования электропотребления

Метрики качества прогноза	Методы прогнозирования			
	<i>Catboost</i>	<i>XGBoost</i>	<i>LGBM</i>	<i>Random Forest</i>
<i>MAPE</i>	2,53%	3,35%	3,23%	3,19%
<i>RMSE</i>	10,79	14,63	13,75	13,63
<i>R2</i>	0,97	0,94	0,95	0,95

Значения метрик качества модели (в табл. 4) указывают на то, что модель метода Catboost является лучшей из представленных в исследовании. Так, метрика RMSE (корень из среднеквадратической ошибки), в которой используются те же единицы измерения, что и в исходных данных, указывает, что разница между фактическими и прогнозируемыми значениями невелика и составляет 10.79 МВт\*ч. Коэффициент детерминации  $R^2$  объясняет, насколько модель соответствует фактическим данным в задачах регрессии. Значение  $R^2 = 0,97$  (близко к 1), т. е. 97 % вариативности в данных объясняется моделью. Также для оценивания качества прогноза использовалась метрика MAPE, которая выражает среднюю абсолютную процентную разницу между фактическими и прогнозируемыми значениями. Метрика, характеризующая среднее отклонение модели от фактических значений,  $MAPE = 2,53 \%$  (хороший результат).

На основании полученных результатов сделан вывод о том, что для краткосрочного (на сутки вперед) прогнозирования электропотребления целесообразно использовать метод машинного обучения Catboost, так как он лучше выявляет существующие закономерности и зависимости в данных.

### 3. Результаты

Проведённые исследования позволили установить, что использование методов машинного обучения на основе градиентного бустинга деревьев решений для прогнозирования электропотребления на краткосрочный период является эффективным подходом.

Методы градиентного бустинга деревьев решений, использованные в исследовании, показывают хороший результат, уменьшают дисперсию прогнозов и обладают устойчивостью к переобучению, что доказывает их функциональность на больших обучающих выборках.

В дальнейшем планируется исследовать возможность синтеза представленных методов машинного обучения для возможности повышения точности прогнозирования.

### Литература

1. Горшенин А.Ю., Денисова Л.А. Прогнозирование выработки электроэнергии ветроэлектростанцией с применением рекуррентной нейронной сети // Известия Тульского государственного университета. Технические науки. 2023. № 4. С. 39–45.
2. Горшенин А.Ю., Васина Д.И. Сравнение используемых методов при прогнозировании выработки электроэнергии ветроэлектростанциями // Математические структуры и моделирование. 2023. № 3 (67). С. 36–42. DOI: 10.24147/2222-8772.2023.3.36-42.
3. Васина Д.И. Актуальные проблемы применения солнечной электроэнергии и прогнозирования // Информационные технологии в науке и производстве : Материалы X Всероссийской молодёжной научно-технической конференции (Омск, 18 апреля 2023 г.) / редколлегия: А.Г. Янишевская (отв. ред.) [и др.]. Омск : Омский государственный технический университет, 2023. С. 89–91.
4. Васина Д.И. Описание программы для проведения анализа выработки мощности постоянного и переменного токов солнечной электростанции // Актуальные вопросы энергетики : материалы Всероссийской научно-практической конференции с международным



- участием (Омск, 25–26 мая 2023 г.) / редколлегия: П.А. Батраков (отв. ред.) [и др.]. Омск: Омский государственный технический университет, 2023. С. 133–137.
5. Гаак В.К., Гаак А.В., Грицай А.С. Повышение экономической эффективности теплоисточников и систем теплоснабжения в современных условиях // Промышленная энергетика. 2017. № 12. С. 14–20.
  6. Отчёт за 2022 г. // СО ЕЭС: сайт. URL: [https://www.so-ups.ru/fileadmin/files/company/reports/disclosure/2023/ups\\_rep2022.pdf](https://www.so-ups.ru/fileadmin/files/company/reports/disclosure/2023/ups_rep2022.pdf) (дата обращения: 02.09.2023).
  7. Горшенин А.Ю. Формирование выборки исходных данных для машинного обучения модели краткосрочного прогнозирования электропотребления // Автоматизация в промышленности. 2023. № 10. С. 37–41. DOI: 10.25728/avtprom.2023.10.08.
  8. Горшенин А.Ю., Васина Д.И. Значение анализа данных в прогнозировании ветроэнергетики // Россия молодая: передовые технологии – в промышленность. 2023. № 1. С. 61–65. DOI: 10.25206/2310-4597-2023-1-61-65.
  9. Горшенин А.Ю. Анализ параметров ветроэлектростанций как объекта управления // Прикладная математика и фундаментальная информатика : материалы XII Международной молодежной научно-практической конференции с элементами научной школы (Омск, 16–21 мая 2022 г.) / отв. ред. А.В. Зыкина. Омск : Омский государственный технический университет, 2022. С. 80–81.
  10. Tyunkov D.A. [et al.] Short-term forecast methods of electricity generation by solar power plants and its classification // Journal of Physics: Conference Series. 2019. Vol. 1260, 5. P. 052033. DOI: 10.1088/1742-6596/1260/5/052033.
  11. Серебряков Н.А. Применение адаптивного ансамблевого нейросетевого метода для краткосрочного прогнозирования электропотребления электротехнического комплекса районных электрических сетей // Омский научный вестник. 2021. № 1 (175). С. 39–45.
  12. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: unbiased boosting with categorical features // Advances in Neural Information Processing Systems, 2018. P. 6638–6648.
  13. Chen T., Guestrin C. XGBoost: A scalable tree boosting system. 2016. arXiv.org. DOI: 10.48550/arXiv.1603.02754.
  14. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree // Neural Information Processing Systems 30 (NIPS 2017). P. 3149–3157.
  15. Ho T.K. Random Decision Forests // Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. P. 278–282.
  16. Потапов В.И., Грицай А.С., Тюньков Д.А. Спектральный анализ ретроспективных данных ООО «Омская энергосбытовая компания» об электропотреблении // Омский научный вестник. 2016. № 5 (149). С. 74–76.
  17. Шугунов Л.Ж., Куповых Г.В. Разложение, анализ и прогноз временных рядов метеопараметров // Известия ЮФУ. Технические науки. 2005. № 11. С.15–21.
  18. Мусатов М.В., Львов А.А. Анализ моделей метода наименьших квадратов и методов получения оценок // Вестник СГТУ. 2009. № 2 (43). С. 137–140.

**COMPARISON OF METHODS BASED ON DECISION TREES IN THE PROBLEM OF  
SHORT-TERM FORECASTING OF POWER CONSUMPTION**

**A.Yu. Gorshenin**

Postgraduate Student, e-mail: augomgtu@gmail.com

Omsk State Technical University, Omsk, Russia

**Abstract.** The paper examines the issues of short-term forecasting of power consumption using machine learning methods based on decision trees. A spectral analysis was carried out and the main periods in the data were identified. The effectiveness of machine learning methods was compared by calculating the main metrics of regression forecasting models.

**Keywords:** power consumption forecasting, machine learning, decision trees, spectral analysis.

*Дата поступления в редакцию: 14.11.2023*