

УДК 519.254, 004.852, 620.92

https://doi.org/10.52754/16948645_2023_1_220

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ ПРИ ПЛАНИРОВАНИИ ПОТРЕБЛЕНИЯ ЭНЕРГЕТИЧЕСКИХ РЕСУРСОВ

*Хуснутдинов Александр Олегович, магистрант,
evolext@gmail.com*

*Карманов Виталий Сергеевич, к.т.н.,
karmanov@corp.nstu.ru*

*Новосибирский государственный технический университет,
Новосибирск, Российская Федерация*

Аннотация. В статье рассматривается актуальная проблема прогнозирования потребления энергетических ресурсов, в частности, объемов потребления тепловой энергии для жилых домов. Приведены основные аспекты выбора прогнозной модели в зависимости от постановки задачи и характера прогнозируемых данных. Рассмотрены несколько современных методов прогнозирования многомерных временных рядов и проведено исследование точности прогнозных моделей на реальных данных, также проведено сравнение точности с моделями, основанными на классических статистических методах прогнозирования.

Ключевые слова: прогнозирование, многомерные временные ряды, планирование энергопотребления, машинное обучение, глубокое обучение, интервальное прогнозирование.

APPLICATION OF MACHINE LEARNING METHODS FOR TIME SERIES FORECASTING IN ENERGY RESOURCE CONSUMPTION PLANNING

*Khusnutdinov Alexander Olegovich, Master's student,
evolext@gmail.com*

*Karmanov Vitaly Sergeevich, Candidate of Technical Sciences,
karmanov@corp.nstu.ru*

*Novosibirsk State Technical University,
Novosibirsk, Russian Federation*

Abstract. The article deals with the actual problem of forecasting the consumption of energy resources, in particular, the volume of thermal energy consumption for residential buildings. The main aspects of choosing a forecasting model depending on the formulation of the problem and the nature of the forecasted data are presented. Several modern methods of predicting multidimensional time series are considered and the accuracy of forecast models based on real practical data is studied, and the precision is compared with models based on classical statistical forecasting methods.

Key words: forecasting, multivariate time series, energy consumption planning, machine learning, deep learning, interval forecasting.

1. Введение (Introduction) Топливо-энергетические ресурсы (ТЭР) – это природные и искусственные запасы, используемые для производства и потребления энергии. Традиционно к ним относят: нефть, газ, уголь, воду, а также электрическую и тепловую энергию, получаемые из различных источников. В настоящее время существует несколько причин, по которым необходимо работать в направлении экономии ТЭР, прежде всего это связано с возможностью возникновения мирового энергетического кризиса, вызываемого ограниченностью ресурсов и непрекращающимся ростом потребления энергии, что ведет к нестабильности на энергетическом рынке и росту цен. Кроме того, экономия ТЭР может оказывать влияние на экономические, экологические, политические и другие аспекты жизни социума, например, снижение вредного воздействия на окружающую среду или уменьшения риска конфликтов, связанных с доступом к источникам этих ресурсов.

Прогнозирование объемов потребления энергоресурсов в натуральном выражении является важным инструментом для управления и оптимизации потребления.

Прогнозирование потребления ТЭР позволяет оценить ожидаемый рост или снижение потребления в будущем, что может помочь определить необходимые меры по сокращению потребления или реорганизации производственных процессов. Прогнозирование также помогает в планировании инвестиций в новые источники энергии и создании инфраструктуры для энергетических систем, что в конечном итоге может привести к сокращению потребления энергоресурсов и более эффективному использованию доступных источников энергии.

Цель настоящей работы заключается в рассмотрении современных методов прогнозирования объемов потребления энергоресурсов, основанных на методологии машинного обучения. Данная работа продолжает серию публикаций о результатах исследований в указанной области [1-4].

2. Постановка задачи прогнозирования (Forecasting problem) Динамика потребления ТЭР представляет собой последовательность значений объемов потребления в разные моменты времени, т.е. представляет собой временной ряд $y_0, y_1, \dots, y_t, \dots$, где $y_i \in \mathbb{R}$. Основная цель анализа временного ряда состоит в выявлении закономерностей в его компонентах для составления прогнозной модели:

$$y_{t+d}(\omega) = f_{t,d}(y_0, \dots, y_t; \omega) \quad (1)$$

где y_{t+d} – прогноз модели на шаг d , ω – вектор параметров модели, $f_{t,d}(\cdot; \cdot)$ – прогнозная модель ряда, построенная на основе t наблюдений.

Выбор прогнозной модели временного ряда – сложная и актуальная задача, ее решение требует учета различных факторов. Рассмотрим некоторые из них.

1) Характер прогнозируемых данных: можно выявлять закономерности в предыдущие моменты времени и на их основе строить прогнозы, а можно дополнительно с этим использовать вспомогательные (экзогенные) переменные для учета дополнительных зависимостей. В настоящее время подгон прогнозной модели под одномерный временной ряд перестает быть актуальным – в эпоху больших данных есть возможность дополнительно к исследуемой величине измерять десятки различных характеристик, включая и нечисловые, с различных приборов учета, датчиков и устройств с требуемой периодичностью, получая в результате многомерные временные ряды (МВР). В связи с этим к алгоритмам выдвигается дополнительное требование – как можно более эффективно обрабатывать весь набор данных. Выбор экзогенных переменных является отдельной важной задачей, поскольку добавление одних переменных может привести к повышению точности прогнозирования, а других – к переобучению модели [5].

На выбор прогнозной модели также может повлиять наличие трендовых и сезонных составляющих в данных, некоторые методы прогнозирования способны автоматически выявлять эти закономерности, а другие требуют отдельного покомпонентного рассмотрения.

2) Итоговый результат, который дает модель на выходе в качестве прогноза: это может быть либо конкретные значение в каждый момент времени горизонта планирования (точечное прогнозирование), либо диапазоном значений, которые может принимать исследуемая величина в каждый момент времени горизонта планирования (интервальное прогнозирование) [6], во втором случае модель на вход принимает еще один параметр:

$$y_{t+d}(\omega, \alpha) = f_{t,d}(y_0, \dots, y_t; \omega, \alpha) \quad (2)$$

где α – вероятность реализации прогноза.

3) Продолжительность горизонта планирования: одни модели временных рядов могут составлять точные прогнозы на короткие промежутки и сильно расходятся на последующих шагах, а другие – иметь немного хуже точность на тех же коротких шагах, но сохранять одинаковый уровень точности при прогнозировании на длительные промежутки, выбор модели в этом случае зависит от специфики задачи.

3. Описание методов прогнозирования (Review of forecasting methods) До появления методов машинного и глубокого обучения, в области прогнозирования временных рядов изучались и использовались статистические методы прогнозирования, включающие в себя, например, процессы авторегрессии (AR), процессы скользящей средней (MA) и авторегрессионной скользящей средней (ARMA) и др. [7, 8].

Модель VARMA – это векторная форма модели ARMA, которая учитывает значения и данные об ошибках для нескольких переменных одновременно [9]. VARIMA – наиболее общая модель, является расширением комбинированной модели VARMA для нестационарных временных рядов, алгебраическая форма представления которой может выглядеть следующим образом:

$$Y_t = (I + \Phi_1)Y_{t-1} + (\Phi_2 - \Phi_1)Y_{t-2} + \dots + (\Phi_p - \Phi_{p-1})Y_{t-p} - \Phi_p Y_{t-p-1} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \dots - \Theta_q \varepsilon_{t-q} \quad (3)$$

где Y_t – вектор размерности $(n \times 1)$ значений n переменных в t момент времени, I – единичная матрица размерности n , Φ_i – матрица коэффициентов авторегрессии размерности $(n \times n)$, p – порядок векторной авторегрессии, ε_t – вектор ошибок размерности $(n \times 1)$, Θ_i – матрица коэффициентов скользящего среднего размерности $(n \times n)$, q – порядок векторного скользящего среднего.

Модели AR, MA и их всевозможные модификации по-прежнему обеспечивают точность, близкую к современным моделям машинного обучения, особенно на небольших наборах данных, где вторые работают не наилучшим образом; но, несмотря на простоту этих моделей, они подвержены переобучению.

Среди известных методов машинного обучения, применяемых для прогнозирования временных рядов, выделяется группа методов, основанных на механизме boosting [10], суть которого заключается в агрегировании (ансамблировании) нескольких предсказательных моделей таким образом, чтобы при добавлении новой модели общая ошибка уменьшалась:

$$F_M(x) = \sum_{m=1}^M b_m h(x; a_m) \quad (4)$$

где $F_M(x)$ – объединенная модель из M базовых моделей, b_m – весовой коэффициент базовой модели, $h(x; a_m)$ – базовая предсказательная модель, характеризующаяся некоторым вектором параметров a_m .

В качестве базовой предсказательной модели может быть любая прогнозная модель, но чаще всего объединяют модели, основанные на решающих деревьях из-за простоты их построения [10].

Существует несколько способов добавления моделей в ансамбль, основной из них – gradient boosting, заключается в конструировании модели таким образом, чтобы они были

максимально коррелированы с отрицательным градиентом функции потерь всего ансамбля:

$$F_M = F_{M-1} - b_M \nabla Q, \quad (5)$$

где Q – вещественная функция, имеющая вид:

$$Q = \sum_{i=1}^N L(y_i, F_M(x_i)), \quad (6)$$

где $L(y_i, F_M(x_i))$ $i = \overline{1, N}$ – функция потерь, N – размер набора данных. Функция потерь позволяет количественно измерить, насколько предсказанный ответ $F_M(x_i)$ отличается от истинного значения y_i .

На данный момент существует несколько реализаций механизма gradient boosting: Extreme Boosting (XGBoost), Light Gradient Boosting (LightGBM) и CatBoost [11].

Другой класс современных методов прогнозирования МВР основан на моделях глубокого обучения. Рекуррентные нейронные сети (RNN), такие как Elman network [12] и Long Short-Term Memory (LSTM) [13], были первыми моделями для прогнозирования временных рядов, однако в настоящее время все активнее используются другие архитектуры, использующие в своей основе сверточные нейронные сети (CNN) и Transformers.

Transformers является структурой нейронной сети, которая использует механизм Attention [14] для обработки задач машинного перевода и позволяет модели использовать информацию из других слов во входной последовательности. Multi-Head Attention (МНА) – это метод, встроенный в структуру Transformers, позволяющий модели одновременно сфокусироваться на нескольких аспектах входной последовательности, что помогает создавать более точные прогнозы. МНА является ключевым фактором в достижении новых высот моделями Transformers в задачах обработки естественного языка и последовательностей, таких как временные ряды. Архитектура Temporal Fusion Transformers (TFT) [15], которая была изначально разработана для создания интервальных прогнозов многомерных временных рядов, объединяет собой блоки LSTM и механизм Transformers, что позволяет ей быть одной из самых производительных в настоящее время.

Архитектура Temporal Convolution Networks (TCN) [16] – это архитектура нейронных сетей, используемых для анализа временных данных, таких как речь, музыка, финансовые данные и другие временные ряды. TCN использует в своей основе блоки сверточных слоев, которые обрабатывают фрагменты временного ряда как изображение и позволяют выявлять зависимости между разными переменными, тренды и циклы. Блоки свертки ранее применялись только в задачах компьютерного зрения (CV), за счет их применения можно распараллеливать вычисления и получать быструю производительную модель, обученную на больших наборах данных.

4. Способ исследования (Research method) Для проведения исследования точности и выбора наилучшей прогнозной модели рассмотрим следующие: TFT, TCN, gradient boosting в реализации CatBoost, а также сравним их с классической моделью VARIMA в реализации AutoARIMA [17], позволяющей выбирать автоматически коэффициенты модели под указанные данные.

Подбор весовых коэффициентов современных моделей (TFT, TCN, CatBoost) будет проводиться по схеме, представленной на Рис. 1

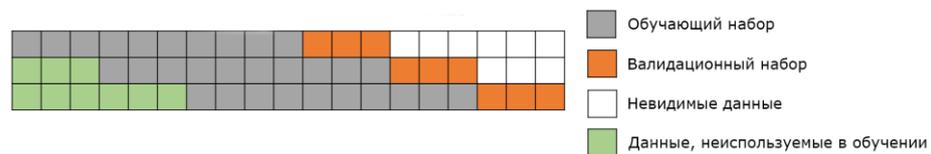


Рисунок 1. Схема валидации моделей при обучении, где одна клетка соответствует одному наблюдению

Подбор оптимальных коэффициентов модели AutoARIMA будет осуществляться сразу на обучающем и валидационном наборе.

Для определения наилучшей из возможных конфигураций каждой из современных моделей, будем варьировать их основные гиперпараметры и определять точность прогнозирования на одних и тех же тестовых данных. Схема определения точности прогнозной модели на тестовых данных представлена на Рис. 2.

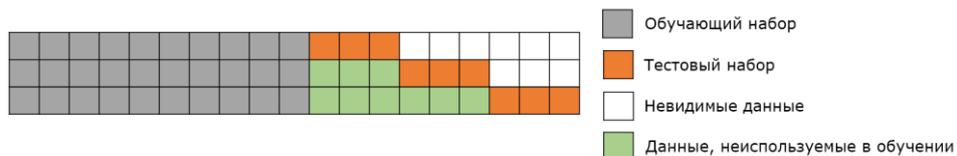


Рисунок 2. Схема определения точности модели на тестовых данных, где одна клетка соответствует одному наблюдению

Кроме варьирования гиперпараметров, будем также изменять число последних наблюдений, используемых современными моделями для составления прогноза – это поможет выявить, как меняется точность модели при уменьшении или при увеличении числа входных данных. Таким образом мы сможем сравнить множество реализаций моделей TFT, TCN и CatBoost с оптимальными весовыми коэффициентами и классической моделью AutoARIMA.

Определять точность прогнозирования моделей будем по значению метрики «Среднеквадратическая ошибка», Root Mean Squared Error (RMSE):

$$RMSE(y, \hat{y}) = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}}, \quad (7)$$

а также по значению метрики «Средняя абсолютная ошибка», Mean Absolute Error (MAE):

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (8)$$

где n – число точек тестового периода, y – истинные значения временного ряда, \hat{y} – прогнозные значения модели.

Также, для тех моделей, которые поддерживают прогнозирование диапазонов, в качестве третьей метрики выберем процент покрытия истинных значений временного ряда доверительным интервалом, образованным квантилями уровня 0.05 и 0.95, метрики RMSE и MAE для таких моделей будут вычисляться для медиан от интервальных прогнозных значений.

5. Используемые данные (Data used) Большая часть актуальных исследований, связанных с прогнозированием МВР, использует данные по потреблению электрической энергии как основу для обучения и определения точности моделей. Причиной этого является большая доступность таких данных с разной периодичностью (ежедневные, почасовые и др.), набором вспомогательных характеристик и продолжительностью. В то же

время исследования по прогнозированию других, например, по потреблению тепловой энергии, практически не приводятся, поскольку таких данных в открытом доступе значительно меньше, они меньше по объему и практически не содержат значения вспомогательных характеристик, изменяющихся во времени. Редко они дополнительно содержат метеорологические параметры: температура наружного и внутреннего воздуха, атмосферное давление, влажность и т.п., а также статические характеристики (площадь помещения, состав оборудования, материалы стен и др.), которые являются малополезными для прогнозирования.

Однако прогнозирование потребления других ТЭР также остается актуальным. В качестве данных для исследования был выбран набор данных с информацией по потреблению тепловой энергии жилыми домами города Томска (Россия). Данные размещены в открытом доступе на портале IEEE DataPort [18] с лицензией Creative Commons Attribution, позволяющей использовать их в подобного рода исследованиях при указании источника.

Данные в наборе представляют собой ежедневные значения потребления с прибора учета тепловой энергии (в Гкал) для жилых домов различной конфигурации за 4 неполных отопительных сезона за период 01/01/2014 по 10/05/2017 гг.

Проведем исследование на одном объекте – жилым девятиэтажным кирпичным доме, данные по потреблению тепловой энергии которого разделим на три набора:

- Обучающий: данные с 01/01/2014 по 30/09/2015, содержит 638 значений;
- Валидационный: данные с 01/10/2015 по 30/09/2016, содержит 342 значения;
- Тестовый: данные с 01/10/2016 по 30/04/2017, содержит 212 значений.

Обучающий и валидационный наборы необходимы для определения оптимальных параметров прогнозных моделей, тестовый – для сравнения качества прогнозирования моделей между собой и определения наилучшей. На Рис. 3 представлена динамика изменения значений потребления для целевого объекта.

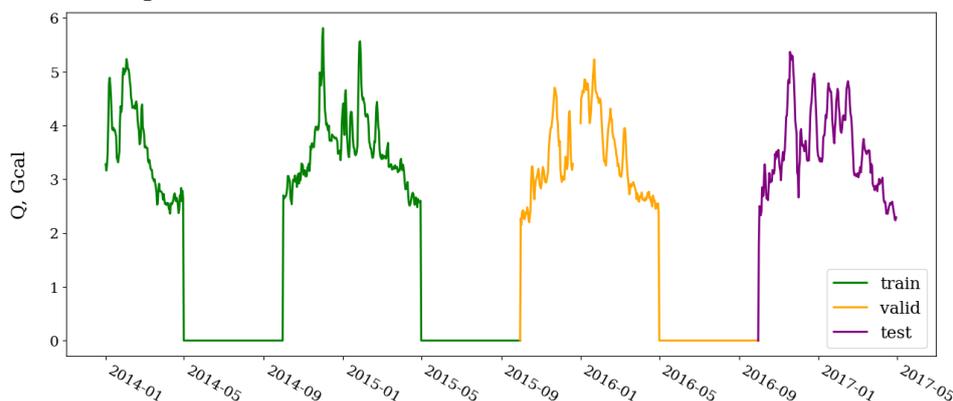


Рисунок 3. Динамика потребления тепловой энергии для целевого объекта, где цветами обозначены обучающая (train), валидационная (valid) и тестовая выборки (test)

Как видно на графике, часть данных валидационного набора отсутствует и, поскольку они содержатся внутри отопительного периода, то в качестве метода восстановления недостающих значений выберем кубический сплайн [19].

Также в качестве внешней учитываемой переменной будем использовать среднесуточную температуру наружного воздуха в соответствующие даты потребления тепловой энергии, на Рис. 4 представлено корреляционное поле внешней и прогнозируемой переменной.

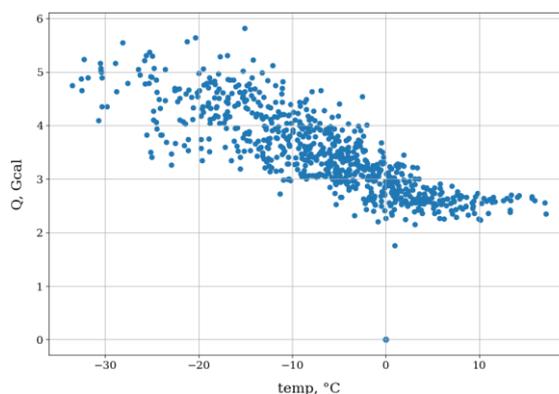


Рисунок 4. Корреляционная карта для температуры наружного воздуха и объема потребления тепловой энергии

Из характера расположения точек на корреляционном поле можно сделать вывод, что между выбранной и целевой переменной существует отрицательная корреляционная зависимость, которая в первом приближении может считаться линейной.

6. Результаты и заключение (Results and conclusion) В Таблице 1 представлены полученные значения метрик на тестовом наборе для выбранных методов при различных значениях гиперпараметров.

Таблица 1. Рассчитанные значения метрик прогнозных моделей на тестовых данных.

Красным и зеленым цветом выделены наихудшие и наилучшие модели интервального прогнозирования соответственно.

Методы и значения гиперпараметров	Метрики	Число входных наблюдений			
		7	15	30	45
AutoARIMA	RMSE, Гкал	2.214			
	MAE, Гкал	2.102			
	Покрытие, %	0			
TCN (kernel_size: 3, num_filters: 3)	RMSE, Гкал	1.266	1.398	1.079	1.141
	MAE, Гкал	0.933	1.277	0.905	0.964
	Покрытие, %	73.11	32.08	60.38	58.49
TCN (kernel_size: 5, num_filters: 5)	RMSE, Гкал	1.972	1.452	1.560	1.571
	MAE, Гкал	1.918	1.335	1.433	1.451
	Покрытие, %	31.60	68.40	57.08	57.08
TFT (dropout: 0.1, full_attention: True)	RMSE, Гкал	0.840	1.116	1.161	1.195
	MAE, Гкал	0.416	1.022	1.071	1.102
	Покрытие, %	81.60	75.94	73.11	71.23
TFT (dropout: 0.25, full_attention: False)	RMSE, Гкал	1.145	1.154	1.241	1.253
	MAE, Гкал	1.049	1.063	1.154	1.165
	Покрытие, %	75.00	75.94	77.36	71.70
CatBoost (l2_lead_reg: 1, depth: 2)	RMSE, Гкал	0.905	0.957	1.214	1.288
	MAE, Гкал	0.503	0.560	0.779	0.827
	Покрытие, %	94.34	85.38	84.43	74.06
CatBoost (l2_lead_reg: 1, depth: 3)	RMSE, Гкал	0.871	0.888	1.105	1.073
	MAE, Гкал	0.497	0.516	0.711	0.708
	Покрытие, %	93.87	91.51	84.43	90.57
CatBoost (l2_lead_reg: 2, depth: 2)	RMSE, Гкал	0.994	0.918	1.166	1.115
	MAE, Гкал	0.575	0.520	0.749	0.722

	Покрытие, %	91.04	93.40	86.79	76.42
--	-------------	-------	-------	-------	-------

Анализ результатов показывает, что классическая модель ARIMA в реализации AutoARIMA показывает наихудшие результаты среди рассмотренных моделей, к тому же совсем не покрывает прогнозируемым доверительным интервалом истинные значения ряда. Наихудшие результаты современными моделями получены для TCN, как с точки зрения точности прогноза, так и с точки зрения стабильности результатов: при одних и тех же гиперпараметрах, процент покрытия моделью истинных значений ряда может изменяться от 32.08% до 73.11%. Изменение числа входных данных меньше всего влияет на модель TFT, а наилучшие результаты прогнозирования получены моделью CatBoost при значениях гиперпараметров $l2_lead_reg=1$ и $depth=2$ [20] и использующая 7 последних наблюдений на входе, для которой процент покрытия в наилучшем случае составляет 94.34%, данный результат может быть улучшен добавлением большего числа внешних переменных в набор данных, а наихудший – не менее 80%, что значительно лучше современных моделей глубокого обучения TCN и TFT.

На Рис. 5 наглядно представлены истинные значения тестового периода и прогнозные значения оптимальной модели CatBoost в виде доверительного интервала и его медианы.

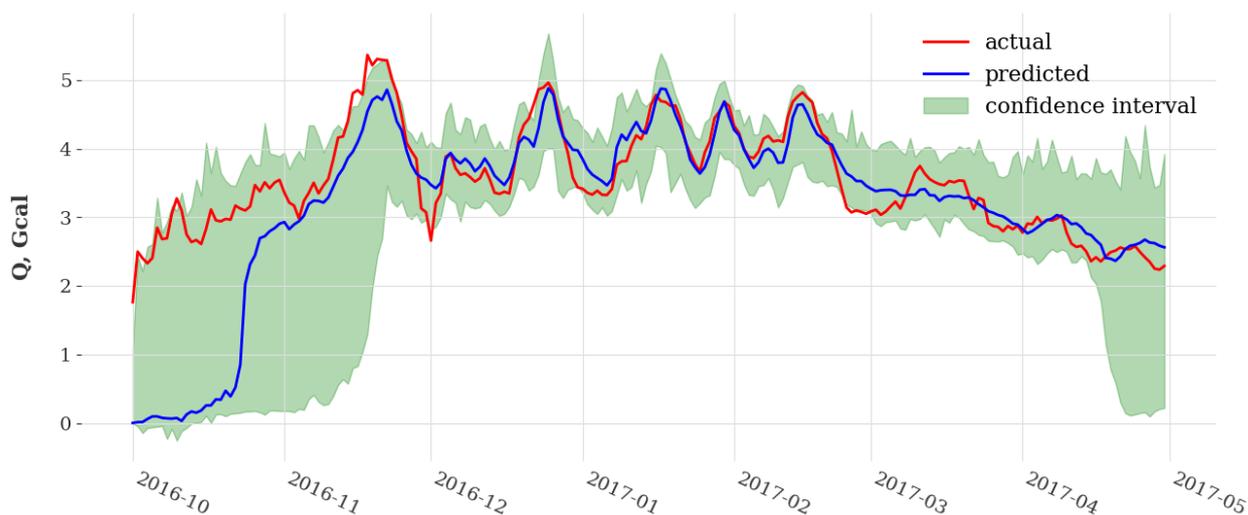


Рисунок 5. Сравнение фактических значений тестового периода и прогнозов модели CatBoost. Цветами выделены фактические (actual), прогнозные (predicted) значения и доверительный интервал (confidence interval)

В работе была рассмотрена задача прогнозирования объемов потребления ТЭР на продолжительные горизонты планирования, в частности, прогнозирование современными методами глубокого и машинного обучения. В работе приведены исследования точности прогнозных моделей на реальных данных при различных настройках и размерах входных данных.

Разработанный подход и полученные оптимальные модели можно рекомендовать к использованию в реальных практических задачах для прогнозирования объемов потребления ТЭР при небольших объемах обучающих данных.

Литература

1. Manusov, V. Analysis of electricity consumption forecasting methods for the coal industry [Text]/ V. Manusov, D. Orlov, V. Karmanov [et al.] // Przegląd Elektrotechniczny. –

2022. – Vol. 98, iss. 9. – P. 26-31. DOI 10.15199/48.2022.09.05

2. Manusov, V. Forecasting Electricity Consumption of Electrical Machines of a Coal Industry Enterprise Using the Wavelet Transform [Text]/ V. Manusov, D. Orlov, V. Karmanov [et. al.]. // IEEE 23 International Conference of Young Professionals in Electron Devices and Materials (EDM) to the 100th anniversary of the legendary NETI rector Georgy Lyshchinsky : IEEE, 2022. – 4 p. – ISBN 978-1-6654-9804-3. DOI 10.1109/EDM55285.2022.9855175

3. Manusov, V. Investigation of Load Schedules of Electrical Machines of a Mining Enterprise Using Wavelet Analysis [Text]/ V. Z. Manusov, D. V. Orlov, P. V. Matrenin [et al.] // IEEE 23 International Conference of Young Professionals in Electron Devices and Materials (EDM) to the 100th anniversary of the legendary NETI rector Georgy Lyshchinsky : IEEE, 2022. – 4 p. – ISBN 978-1-6654-9804-3. DOI: 10.1109/EDM55285.2022.9855045

4. Manusov, V. Predictive Control and Production Process Forecasting Under Deterministic Chaos [Text]/ V. Manusov, D. Orlov, V. Karmanov [et. al.] // IEEE 23 International Conference of Young Professionals in Electron Devices and Materials (EDM) to the 100th anniversary of the legendary NETI rector Georgy Lyshchinsky : IEEE, 2022. – 5 p. – ISBN 978-1-6654-9804-3. DOI: 10.1109/EDM55285.2022.9855192

5. Hastie, T. The Elements of Statistical Learning [Text]/T. Hastie, R. Tibshirani, J. Friedman. – Springer, 2017. – 763 p.

6. Brownlee, J. Introduction to Time Series Forecasting with Python [Text]/ J. Brownlee. – Machine learning mastery, 2020. – 364 p.

7. Manusov, V. Analysis of methods of electricity consumption forecasting for a coal industry enterprise [Text]/V. Manusov, D. Orlov, V. Karmanov [et al.] // IOP Conference Series: Earth and Environmental Science. – 2022. – Vol. 1045, iss. 1 : Ensuring sustainable development in the context of agriculture, green energy, ecology and earth science (ESDCA–2022): 2 International scientific and practical conference, Smolensk, 23–27 Jan. 2022. – Art. 012035 (11 p.). DOI 10.1088/1755-1315/1045/1/012035.

8. Айвазян, С. А. Прикладная статистика. Основы эконометрики : учебник для вузов : в 2-х т. [Text]/ С. А. Айвазян. – М.: ЮНИТИ-ДАНА, 2001. – Т. 2. – 432 с.

9. Rusyana, A. Application of Clustering and VARIMA for Rainfall Prediction [Text]/A. Rusyana, N. Tatsara // IOP Conference Series Materials Science and Engineering. 2020. 796(1):012063. DOI:10.1088/1757-899X/796/1/012063

10. Natekin, A. Gradient boosting machines, a tutorial [Text]/ A. Natekin, A. Knoll // Front. Neurorobot. 2013. Vol. 7. № 21. DOI: 10.3389/fnbot.2013.00021

11. Anghel, A. Benchmarking and optimization of gradient boosting decision tree algorithms [Text]/ A. Anghel, N. Papandreou, T. Parnell [et al.] – URL: <https://arxiv.org/abs/1809.04559> (дата обращения – 30.03.2023 г.)

12. Elman, J. Finding Structure in Time [Text]/ J. Elman // Cognitive Science. 1990. 179-211 pp.

13. Hochreiter, S. Long Short-Term Memory [Text]/S. Hochreiter, J. Schmidhuber // Neural Computation. 1997. Vol. 9. № 8. 1735-1780 pp. DOI: 10.1162/neco.1997.9.8.1735

14. Vaswani, A. Attention Is All You Need [Text]/ A. Vaswani, N. Shazeer, N. Parmar [et al.] – URL: <https://arxiv.org/abs/1706.03762> (дата обращения – 30.03.2023 г.)

15. Lim, B. Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting [Text]/B. Lim, S. Arik, N. Loeff [et al.] – URL: <https://arxiv.org/abs/1912.09363> (дата обращения – 30.03.2023 г.)

16. Bai, S. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling [Text]/ S. Bai, J. Kolter, V. Koltun – URL: <https://arxiv.org/abs/1803.01271> (дата обращения – 30.03.2023 г.)

17. Сапрыкин, К. Автоматизация процесса анализа временных рядов с использованием модели AUTO.ARIMA в R [Текст]/ К. Сапрыкин // Научный журнал. 2019. №5 (39). URL: <https://cyberleninka.ru/article/n/avtomatizatsiya-protsesta-analiza-vremennyh-ryadov-s-ispolzovaniem-modeli-auto-arima-v-r> (дата обращения: 30.03.2023 г.)

18. Zorin, P. Data of heating meters from residential buildings in Tomsk (Russia) for statistical modeling of the thermal characteristics of buildings [Text]/ P. Zorin, O. Stukach – URL: <https://ieee-dataport.org/documents/data-heating-meters-residential-buildings-tomsk-russia-statistical-modeling-thermal> (дата обращения – 30.03.2023 г.)
19. Biloš, M. Irregularly-Sampled Time Series Modeling with Spline Networks[Text]/M. Biloš, E. Ramneantu, S. Günnemann– URL: <https://arxiv.org/abs/2210.10630> (дата обращения – 30.03.2023 г.)
20. CatBoost [Electronic resource] / URL: <https://catboost.ai/docs/> (date of the application: 01.03.2023).