

УДК 621.311

ПРИМЕНЕНИЕ АЛГОРИТМА ГРАДИЕНТНОГО БУСТИНГА ДЛЯ ПРОГНОЗИРОВАНИЯ ЭНЕРГОПОТРЕБЛЕНИЯ СОБСТВЕННЫХ НУЖД ЭЛЕКТРИЧЕСКОЙ СТАНЦИИ

Фатин М.А., студент гр. ЭПм-241 (КузГТУ)

Научный руководитель Паскарь И.Н., ст. преподаватель (КузГТУ)
г. Кемерово

В настоящее время в связи с удорожанием топливных носителей и модернизацией оборудования электрических станций, переход на новые более мощные электроприемники, остро встает вопрос оптимизации и повышения эффективности в сфере потребления электроэнергии на собственные нужды станции. Правильно составленный прогноз электропотребления способен не только минимизировать издержки на выработку электрической энергии, но и выявить неэффективные и убыточные узлы в работе станции.

Существует множество подходов к прогнозированию затрат на собственные нужды, от простых блочных методов [1], до создания нейронных сетей. Первые малоприменимы и используются скорее для оценки поскольку имеют низкую точность конечного результата прогноза. Вторые же напротив имеют высокую точность, однако требуют высокого вычислительного ресурса для их реализации.

Немаловажную роль в прогнозировании помимо учета структуры затрат на собственные нужды играет учет косвенных факторов, влияющих на итоговый результат прогноза. Такими являются: социальный потерн и температура окружающей среды. В разрезе прогнозирования социальный потерн рассматривается как: время суток, выходной, рабочий или праздничный день. Это необходимо поскольку помимо самих затрат на работу генерирующего оборудования в структуру собственных нужд входят затраты на кондиционирование или обогрев помещений, то есть в момент, когда на станции присутствуют только лица из числа оперативного персонала ТЭЦ расходы на обеспечение бытовых помещений – ниже.

Основной задачей моего исследования было провести сравнительный анализ между применением алгоритмов прогнозирования через линейную регрессию и градиентного бустинга (XGBoost). По своей структуре в основе линейной регрессии лежит понятие о линейной связи между всеми членами массива. зависимыми и независимыми переменными, именно на основе высокой линейной зависимости такой алгоритм получает усредненный прогноз за предстоящий период, математически такая функция имеет следующее описание:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_k X_k + \varepsilon \quad (1)$$

где Y – зависимая переменная; $X_1, X_2, \dots, X_K$ – независимые переменные; β_0 – коэффициент сдвига; $\beta_1, \beta_2, \dots, \beta_k$ – коэффициенты наклона; ε – ошибка.

Коэффициенты модели определяются методом наименьших квадратов, который минимизирует сумму квадратов отклонений прогнозируемых значений от фактических. Исходя из этого можно сделать вывод о том, что такая функция не может: учесть нелинейные зависимости, не может учитывать временную динамику, не учитывает внешнее влияние температуры на систему, имеет высокий уровень чувствительности к выбросам. Последний пункт самый важный, поскольку функция не умеет фильтровать получаемые данные, а просто работает с готовым массивом, а потому при появлении ошибочных данных или пустот считывает их как норму и использует далее для составления прогноза.

В сравнении с описанным выше методом, градиентный бустинг можно назвать промежуточным звеном между сложно реализуемыми методами прогнозирования с применением нейронных сетей и простейших аналитических методов [3]. Такой метод не требует высокого вычислительного ресурса, не подвержен влиянию выбросов.

Градиентный бустинг (XGBoost) представляет собой ансамблевый метод с применением алгоритмов градиентного спуска и математически имеет следующий вид:

$$\mathcal{L}(\theta) = \sum_{i=1}^n l(y_i \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (2)$$

где $l(y_i \hat{y}_i)$ – функция потерь (квадратичная ошибка), $\Omega(f_k)$ – регуляризационный член для k -го дерева, учитывающий его сложность.

Такой метод не подвержен влиянию выбросов поскольку создает прогноз поэтапно. Производя множество итераций, где каждая последующая попытка учитывает ошибки предыдущего прогноза, становится возможно достичь высокой точности прогноза. Достигается при помощи регуляризационного члена в исходном выражении [3]:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda |w|^2 \quad (3)$$

где T – количество ветвей в дереве решений, w – вес каждой ветви, γ и λ – параметры регуляризации.

Благодаря возможности включить в расчетную модель дополнительные параметры, не имеющие прямой связи с данными из массива для прогнозирования – становится возможно учесть температурные параметры и социальный потерн.

Для подтверждения того, что применение более комплексных методов для прогнозирования способно составить прогноз с более высокой ценностью, мы провели сравнительный анализ результатов, полученных при применении метода простой линейной регрессии и градиентного бустинга, для этого был сформирован обучающий массив данных. Данные об энергопотреблении представлены предприятием по часам, в период с 2019 по 2023 год, такая обучающая выборка является благоприятной средой для прогнозирования различными способами, поскольку массив данных – обширный (Таблица 1).

Таблица 1

Пример обучающей выборки

Год-месяц	tg1 0	tg11	tg1 2	tg1 3	tg3	tg5	tg6	tg7	tg9	Фак потребле ния
2019-01-01 00:00:00.000	1	1	1	0	0	1	0	1	1	43,485
2019-01-01 01:00:00.000	1	1	1	0	0	1	0	1	1	42,186
2019-01-01 02:00:00.000	1	1	1	0	0	1	0	1	1	42,777
2019-01-01 03:00:00.000	1	1	1	0	0	1	0	1	1	43,017
2019-01-01 04:00:00.000	1	1	1	0	0	1	0	1		45,473
2019-01-01 05:00:00.000	1	1	1	0	0	1	0	1	1	46,241
2019-01-01 06:00:00.000	1	1	1	0	0	1	0	1	1	46,121
...	...	...	...	...	...	...	...	...	...	...
2023-11-13 21:00:00.000	0	0	0	0	0	0	0	0	0	38,908
2023-11-13 22:00:00.000	0	0	0	0	0	0	0	0	0	37,992
2023-11-13 23:00:00.000	0	0	0	0	0	0	0	0	0	38,016

Ниже приведен графический и показательный вывод прогноз методом простой линейной регрессии (таблица 2 и рис.1)

Таблица 2

Показательные величины расчета

Показательная величина	Расчетное значение
MSE	26,69
RMSE, МВт	5,17
MAE, МВт	4,23
Коэффициент детерминации R^2	0,78

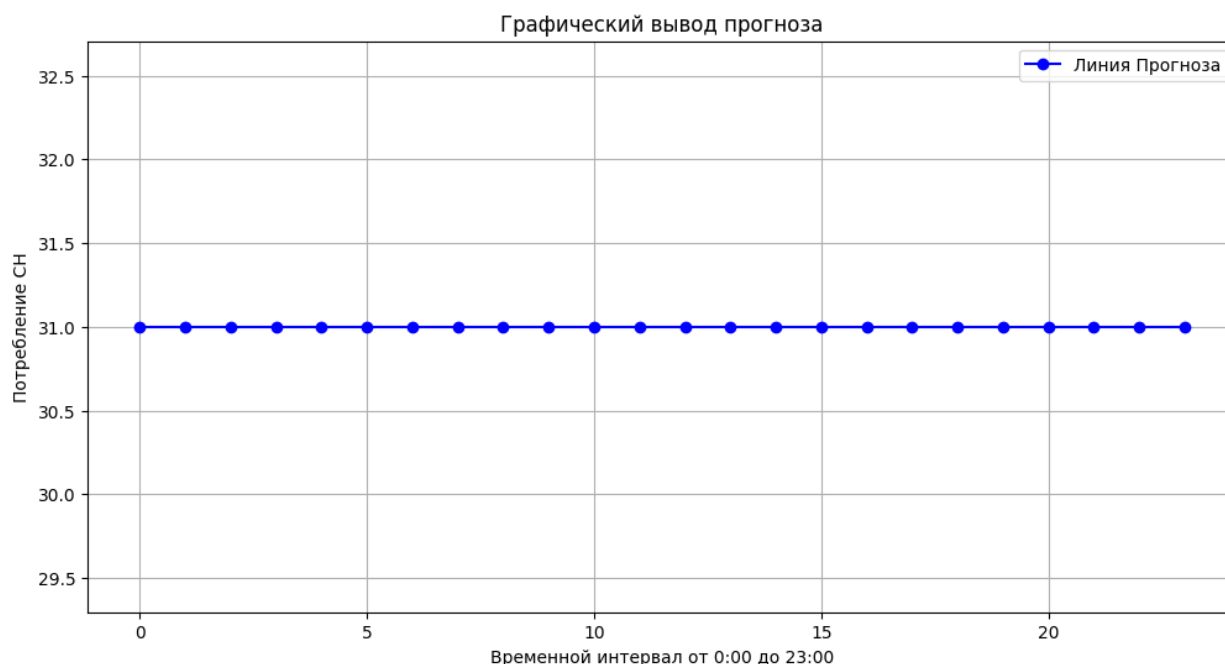


Рисунок 1 Графический вывод результатов прогноза методом линейной регрессии

Анализируя полученные результаты следует дать пояснение для исследуемых метрик прогноза MSE – средняя квадратичная ошибка, определяется как среднее арифметическое квадратов отклонения прогноза от фактических значений. RMSE – оценка отклонения прогноза выраженная в показательной величине, в нашем случае МВт, определяется как квадратный корень из MSE. MAE- средняя абсолютная ошибка, медианной значение абсолютной ошибки, в отличие от MSE менее чувствительно к выбросам поскольку не возводит их в квадрат. Коэффициент детерминации R^2 – показательная величина линейной зависимости между членами массива, где 0 – отсутствие линейной зависимости в массиве данных, а 1 – высокая степень линейной зависимости [3, 4, 5].

Ниже приведены результаты прогноза методом XGBoost (таблица3, рис.2)

Таблица 3

Показательные величины расчета	
Показательная величина	Расчетное значение
MSE	1,83
RMSE, МВт	1,35
MAE, МВт	1,08
Коэффициент детерминации	0,92

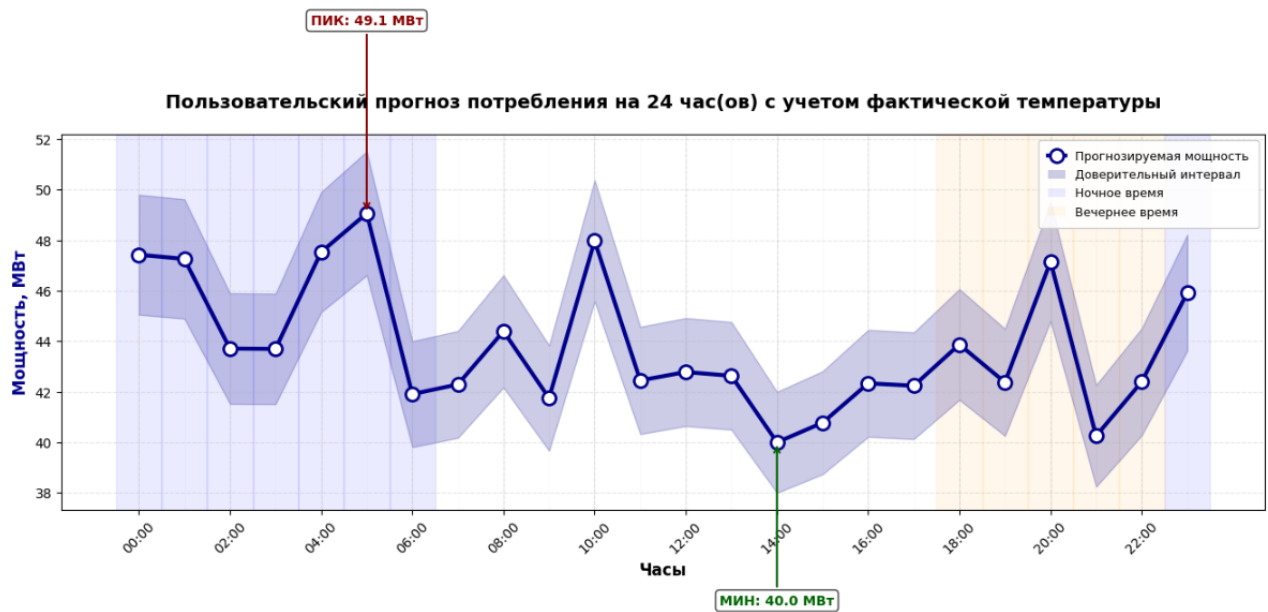


Рисунок 2 Графический вывод результатов прогноза методом XGBoost

Проведя сравнительный анализ, можно отметить, что результаты, получаемые при применении более сложных методов прогнозирования, имеют более высокую точность. Это можно отследить по двум явным показателям, алгоритм градиентного бустинга в сравнении с регрессионным методом добился снижения среднеквадратичной ошибки прогноза почти в 15 раз и смог выявить больше линейных зависимостей из исходного массива данных на 18%.

Такой результат стал возможен благодаря учету следующих метрик при составлении прогноза (таблица 4):

Таблица 4

Дополнительны метрики для прогноз методом градиентного бустинга

Ветвь в дереве решений (T)	Вес ветви дерева решений (w)
Потребление предыдущего часа (инерционность процесса)	0,284
Фактическая температура воздуха	0,156
Циклическая составляющая времени суток	0,098
Потребление в аналогичный час предыдущих суток	0,087
Разброс температур (неравномерность нагрузок)	0,042
Ночное время суток	0,038
Зимний сезон	0,031

На основе проведенного исследования можно сделать вывод о том, что комбинирование методов для подготовки прогнозов и введение дополнительных метрик для расчётов имеют благотворное влияние на конечный получаемый результат. Таким образом применение ансамблевых методов с расширенной библиотекой, исследуемых признаков, существенно повышает точность

получаемых прогнозов и является актуальным для решения задач по оптимизации и снижению потерь на собственные нужды электрической станции.

Список литературы:

1. «Интервальное прогнозирование временных рядов с помощью рекуррентных нейронных сетей с долгой краткосрочной памятью» [Электронный ресурс]: Режим доступа: <https://habr.com/ru/articles/505338/>
2. «Краткосрочное прогнозирование электропотребления сетевой компании» Жилин Б.В. Исаев А.С. Андреев Д.Е. [Электронный ресурс]: Режим доступа: <https://cyberleninka.ru/article/n/kratkosrochnoe-prognozirovanie-elektropotrebleniya-setevoy-kompanii>
3. «Метод градиентного бустинга при прогнозировании управленческих решений в многослойной криогенной системе» Астраханцева И.А. Котене Т.Е. Горев С.В. Астраханцев Р.Г. Грименций П.Н. [Электронный ресурс]: Режим доступа: <https://cyberleninka.ru/article/n/metod-gradientnogo-bustinga-pri-prognozirovanii-upravlencheskih-resheniy-v-mnogosloynoy-kriogennoy-sisteme>
4. «Прогнозирование потребления электрической энергии промышленным предприятием с помощью методов машинного обучения» Моргоева А.Д. Моргоев И.Д. Ключев Р.В. Гаврина О.А. [Электронный ресурс]: Режим доступа: <https://cyberleninka.ru/article/n/prognozirovanie-potrebleniya-elektricheskoy-energii-promyshlennym-predpriyatiem-s-pomoschyu-metodov-mashinnogo-obucheniya>
5. «Прогнозирование потребления электрической энергии промышленным предприятием с помощью методов машинного обучения» Моргоева А.Д. Моргоев И.Д. Ключев Р.В. Гаврина О.А. [Электронный ресурс]: Режим доступа: <https://izvestiya.tpu.ru/archive/article/view/3527>

Информация об авторах:

Фатин Михаил Андреевич, студент гр. ЭПм-241, КузГТУ, ул. Весенняя, д. 28, г. Кемерово, 650000, Россия, fatin_m2020@mail.ru

Паскарь Иван Николаевич, старший преподаватель, КузГТУ, ул. Весенняя, д. 28, г. Кемерово, 650000, Россия, paskar-ivan@mail.ru