

Предсказание потребления электроэнергии

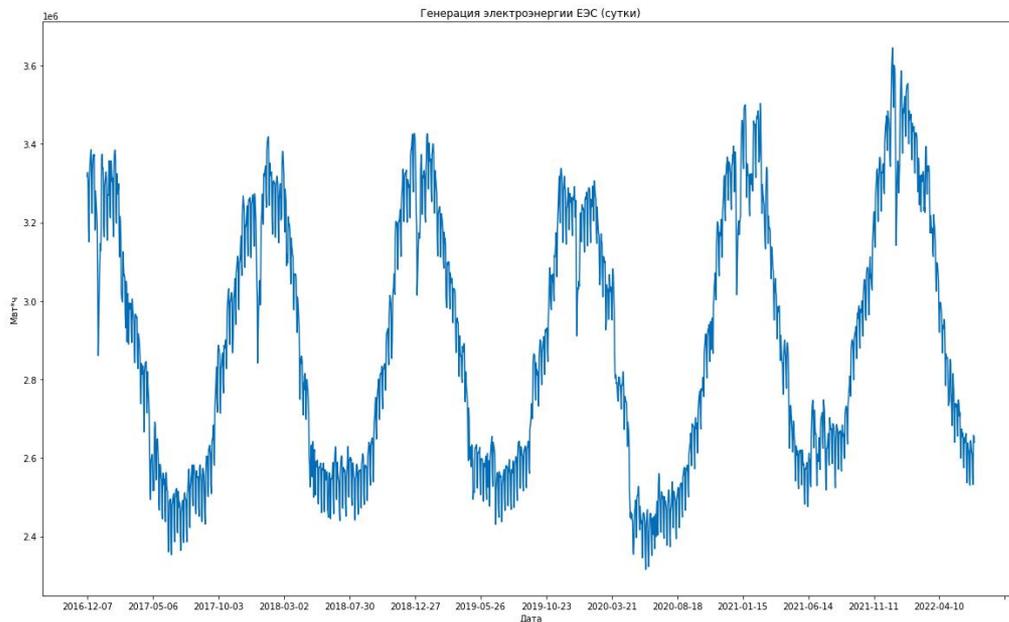
1. Постановка задачи
2. Статистический подход (ARIMA)
3. LSTM
4. NHITS
5. Итоги

Постановка задачи

Задача: Необходимо сформировать прогноз потребления ЭЭ на следующие сутки/7 суток/30 суток на основе исторических данных потребления, температуры и графика рабочих дней

Источники информации:

1. Факт потребления ЭЭ по суткам
2. Фактическая температура наружного воздуха
3. Прогноз температуры наружного воздуха
4. График рабочих и нерабочих дней (также следует учитывать предпраздничные дни)



Статистический подход $ARIMA(p, d, q)$

1. Модель скользящего среднего $MA(q)$
2. Модель авторегрессии $AR(p)$
3. Модель над разностями порядка d

Модель скользящего среднего MA(q)

Модель скользящего среднего анализирует ошибку в предсказаниях в прошлом, для того чтобы улучшить предсказания в настоящем. Уравнение MA имеет следующий вид:

$$Y_t = \beta_2 + \omega_1 \epsilon_{t-1} + \omega_2 \epsilon_{t-1} + \dots + \omega_q \epsilon_{t-q} + \epsilon_t$$

Здесь ϵ это наблюдаемые ошибки в прошлом, а веса ω высчитываются по корреляциям. В данном уравнении q обозначает длину скользящего окна, т.е. количество шагов в прошлом, которые значительно влияют на текущее наблюдение.

Модель авторегрессии AR(p)

Регрессивная модель, в которой зависимая переменная зависит от предыдущих значениях себя. Это означает, что текущие значения коррелируют со значениями в прошлом. Уравнение AR имеет следующий вид:

$$Y_t = \beta_1 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p}$$

Веса Φ_1, \dots, Φ_p относящиеся к значениям в прошлом высчитываются по величине корреляции между текущим значением и предыдущим, так, большей корреляции сопоставляются большие веса.

Модель ARIMA

Суммируя приведенные в предыдущих слайдах уравнения получаем модель ARMA. Для того, чтобы уметь предсказывать тренд, проведем следующую операцию: $Z_t = Y_{t+1} - Y_t$ ее можно повторить d раз, тем самым попытаться избавиться от всех трендом. Модель ARMA, примененная к итоговому ряду называется ARIMA.

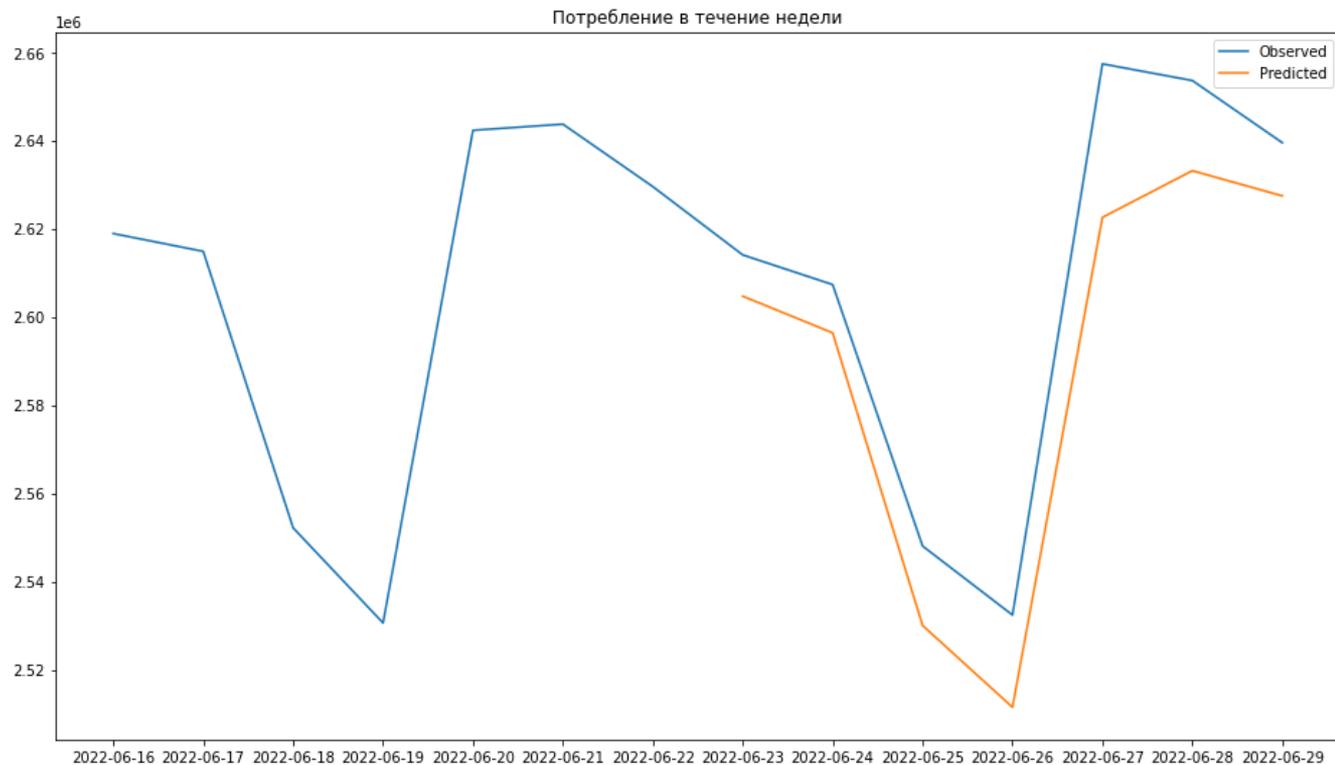
Результаты

В нашем случае используется еще и температура, таким образом к итоговому предсказанию прибавляются значения температуры на предыдущих шагах, с различными весами. Все веса находятся по методу наименьших квадратов.

Итоговая метрика качества:

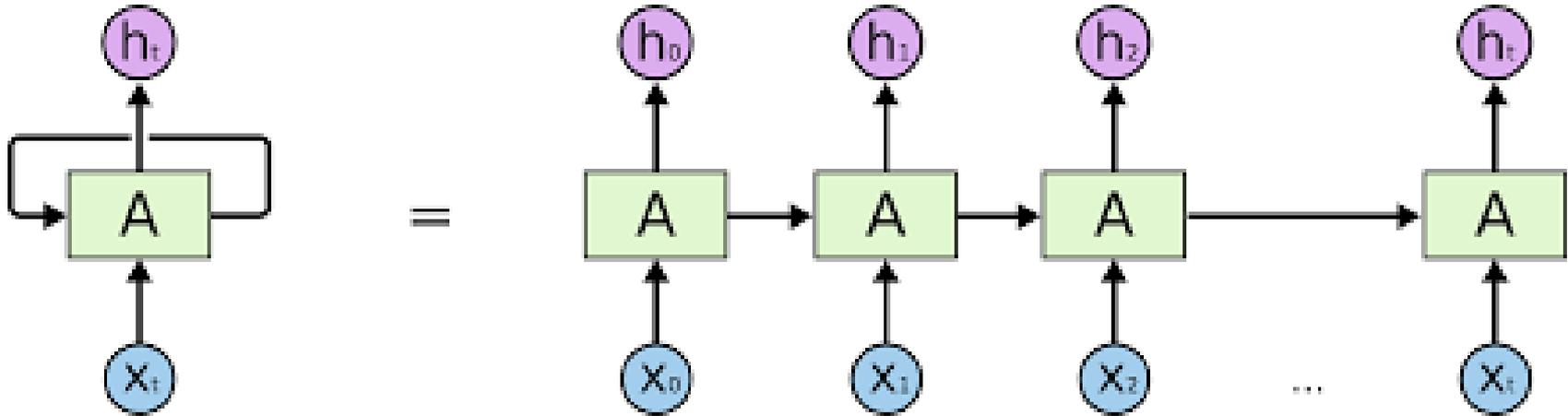
$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Результаты

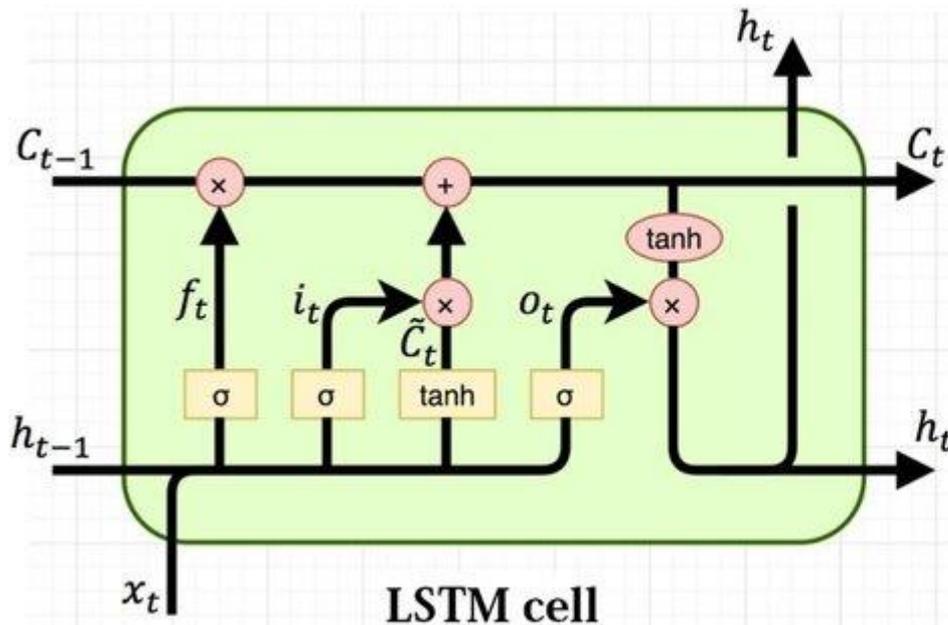


MAPE = 0.7%

LSTM



LSTM



$$i_t = \sigma(x_t U^i + h_{t-1} W^i)$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o)$$

$$\tilde{C}_t = \tanh(x_t U^g + h_{t-1} W^g)$$

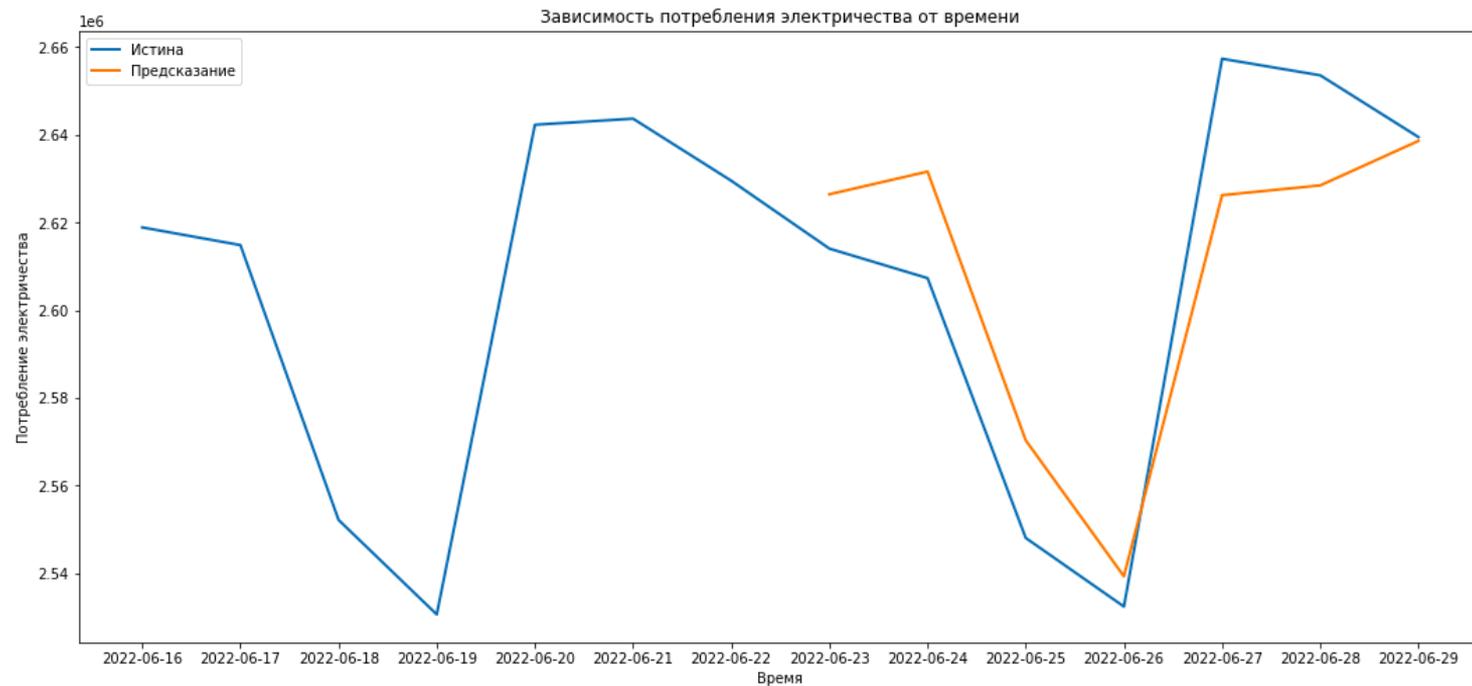
$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t)$$

$$h_t = \tanh(C_t) * o_t$$

LSTM

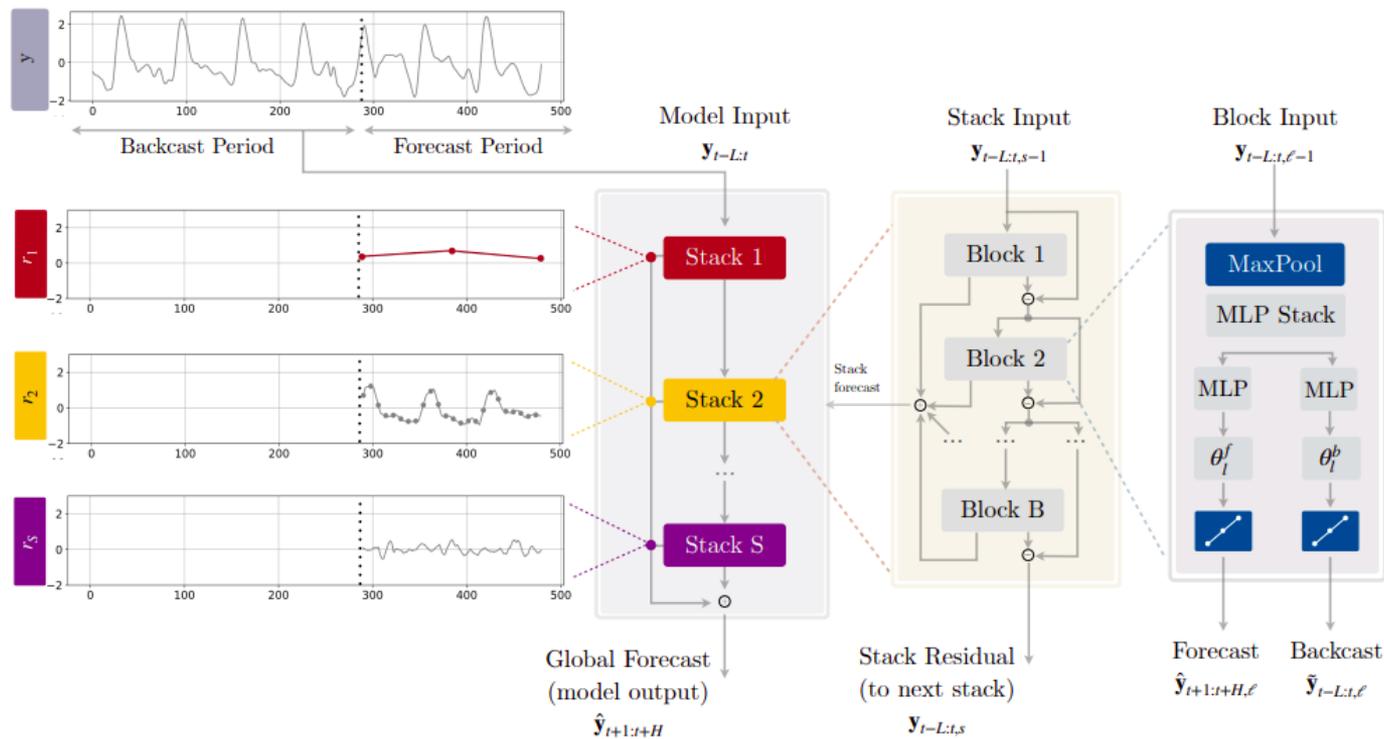
Итоговая модель обучается на данных о потреблении электроэнергии ЕЭС и температуре с 07.12.2016 по 22.06.2022. Также в модель была добавлена информация о выходных днях. Предсказание формируется на одну неделю, основываясь на информации о предыдущей неделе, а также температуре на следующую.

LSTM

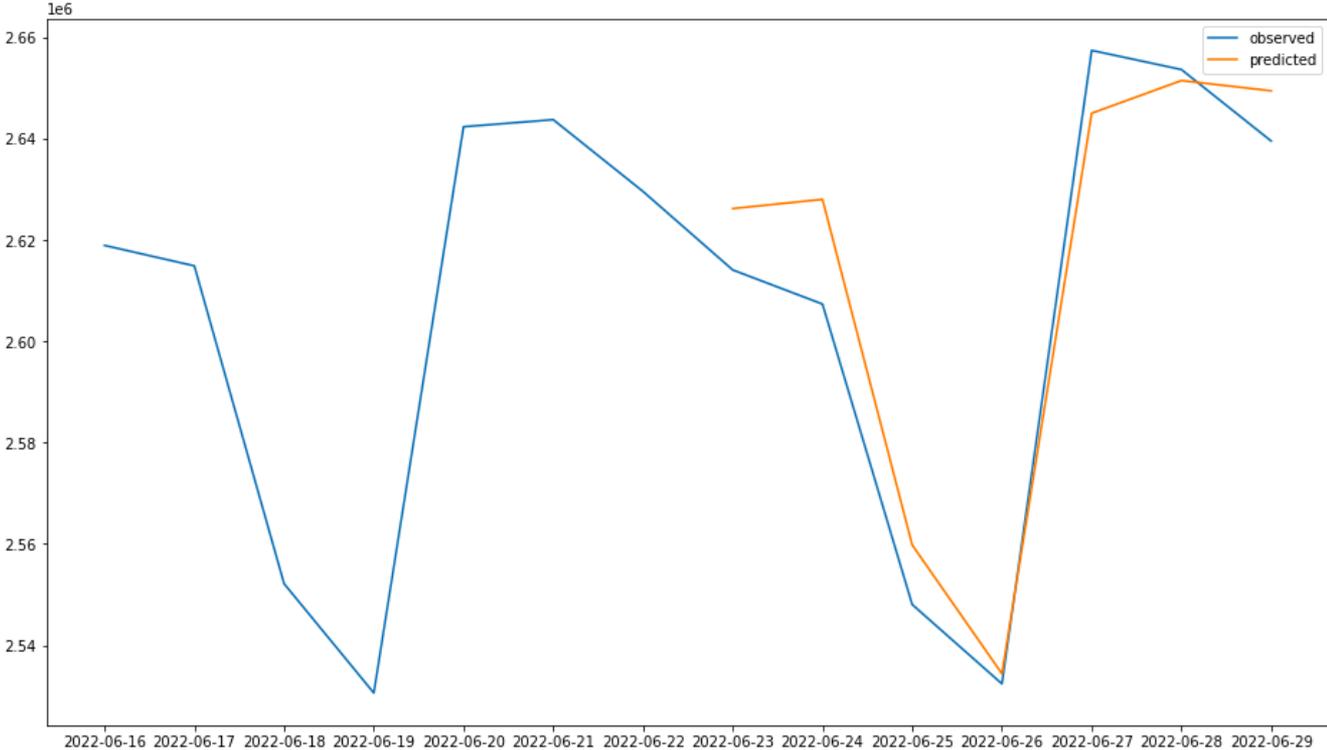


MAPE = 0.67%

NHiTS - Neural Hierarchical Interpolation for Time Series Forecasting

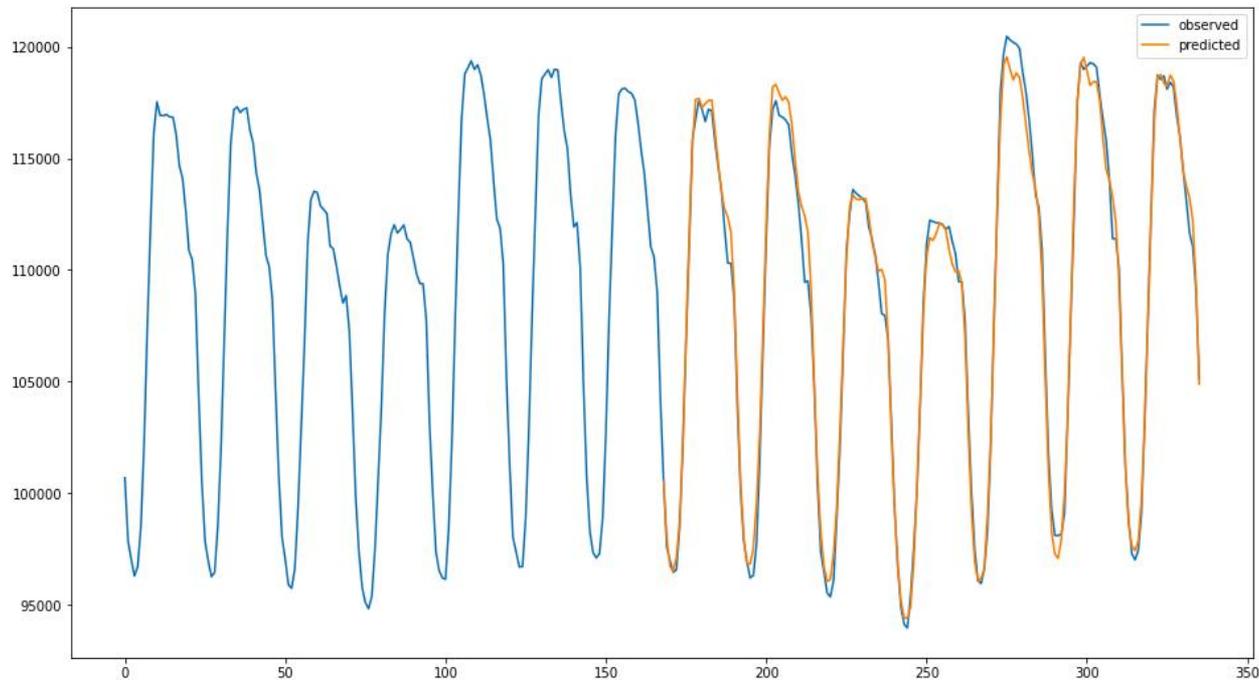


NHiTS



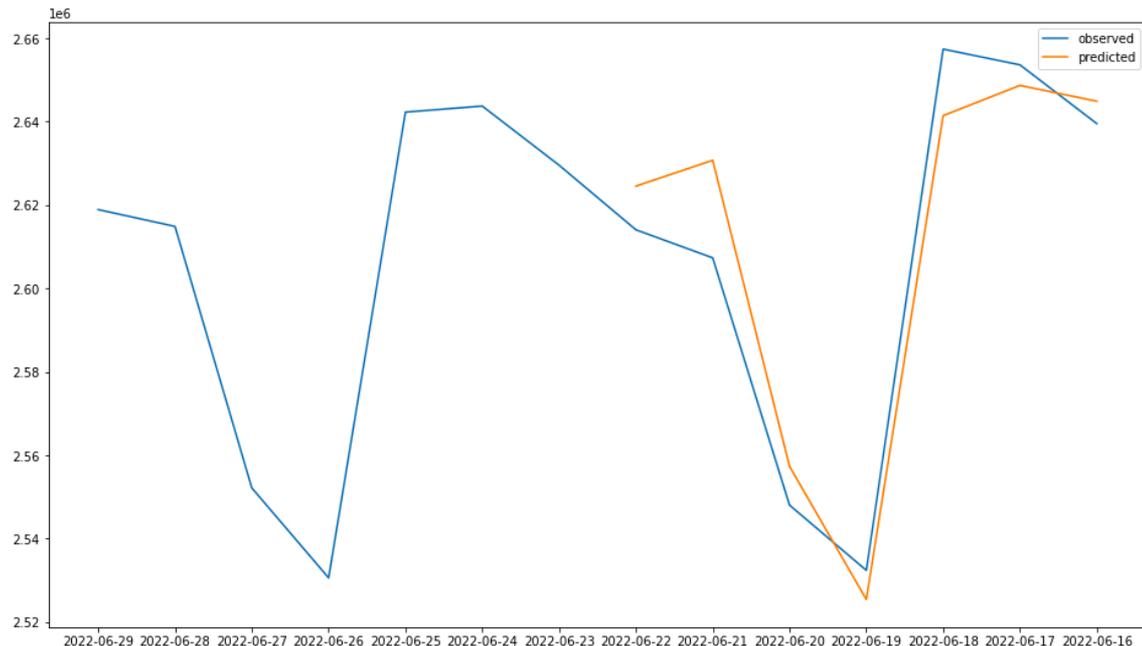
MAPE = 0.39%

Модель с использованием почасовых данных с последующей агрегацией до суток



Предсказание на 7 * 24 часов с 23.06.2022 по 29.06.2022

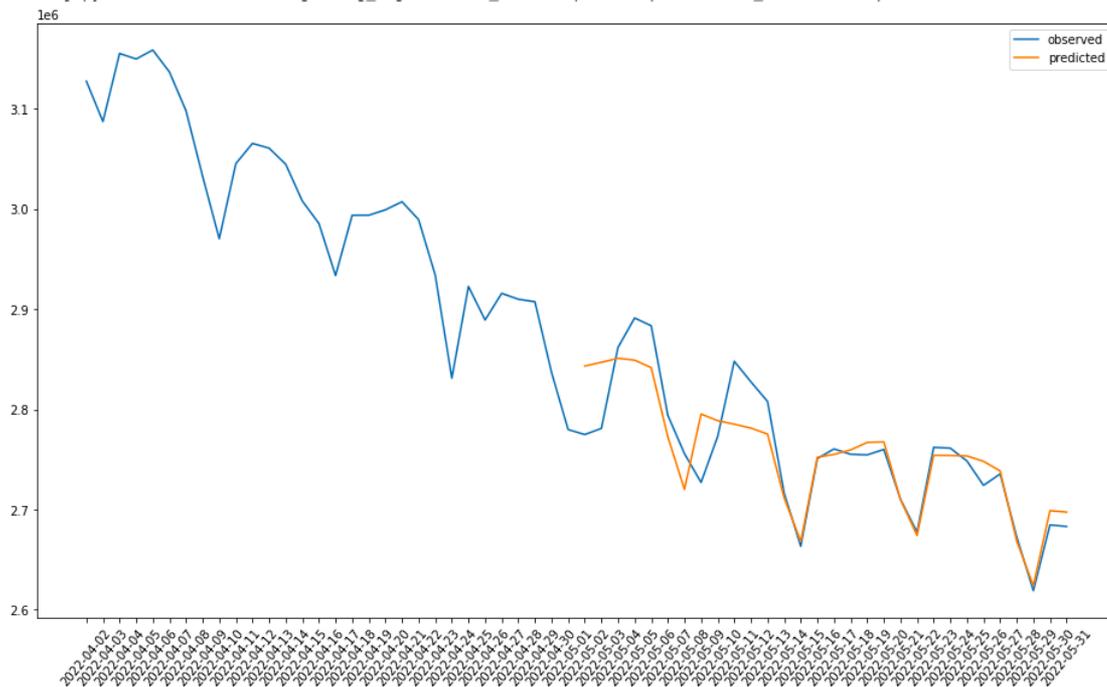
Модель с использованием почасовых данных с последующей агрегацией до суток



Имея предсказание на каждый час, можем просуммировать 24 часа, получив предсказание на день.

MAPE = 0.5%

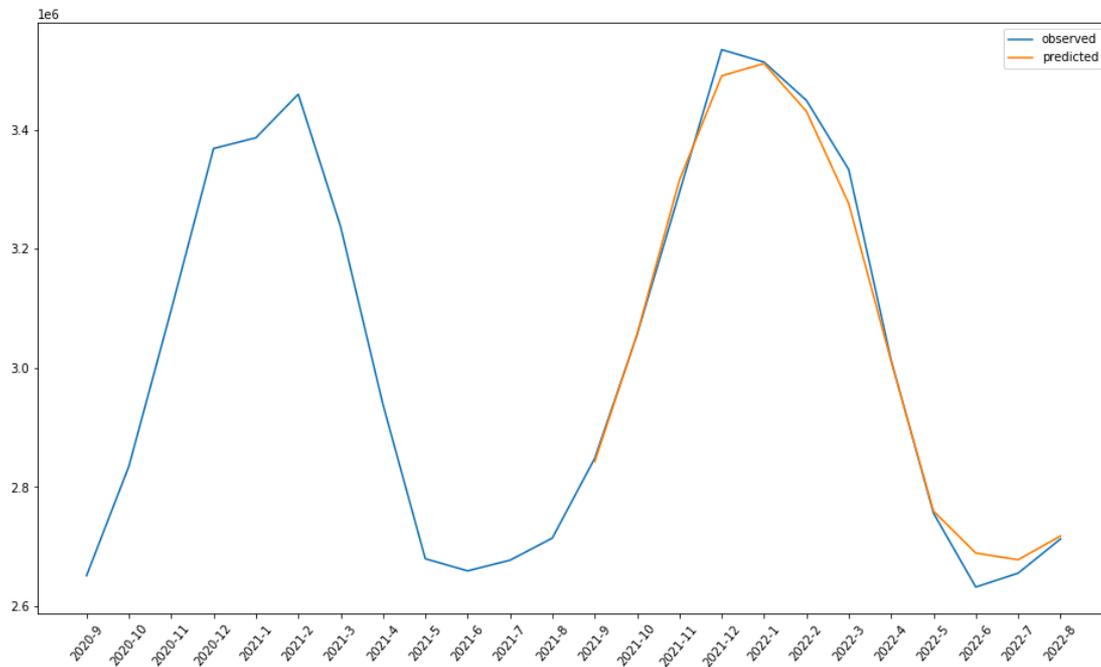
Предсказания на месяц



Для предсказания на месяц размер выхода модели 30 точек. Данное предсказание на май 2022 на 30 дней.

MAPE = 0.77%

Предсказания на год



Для предсказания на год
выход 12 точек.
Потребление энергии на
месяц считается как сумма
потреблений за каждый
день месяца.

MAPE = 0.65%

Итоги

Точность результатов предсказаний на неделю

	ЕЭС, MAPE
ARIMA	0.7%
LSTM	0.67%
NHITS	0.39%