

Энергетический оракул

2 этап

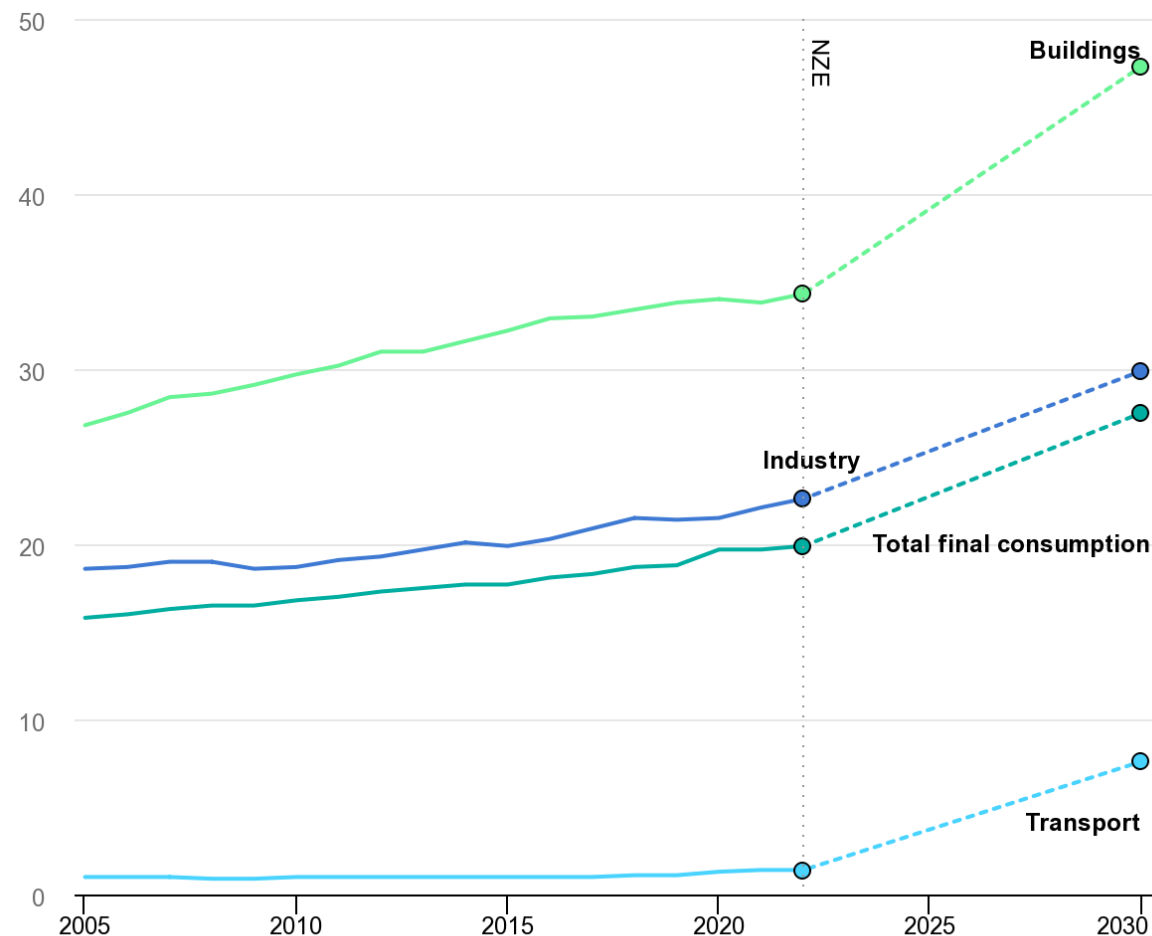
О ЧЕМ ЭТОТ КЕЙС?

Для создания кейса использовались реальные данные. В ходе решения кейса могут использоваться как предоставленные данные, так и данные из открытых источников, найденные в сети Интернет.

Постоянный рост потребления электроэнергии один из наиболее заметных и сложных вызовов, стоящих перед современным обществом. В течение последних десятилетий спрос на электричество стремительно возрастает, и это явление имеет глобальный характер. Повышение потребления электроэнергии обусловлено несколькими факторами, которые оказывают влияние на образ жизни и технологические тенденции.

Факторы влияющие на рост потребления:

- 1.Первым и наиболее значимым фактором является постоянный рост населения во всем мире.
- 2.Вторым фактором, содействующим росту потребления электроэнергии, является технологический прогресс и цифровизация общества.
- 3.Третий фактор связан с расширением применения электроэнергии в промышленности.



ОПИСАНИЕ ПРОБЛЕМЫ

Прогнозирование электропотребления представляет собой сложную задачу из-за нескольких ключевых проблем:

1. Неопределенность: Будущее поведение потребителей и их потребности в электроэнергии не всегда можно предсказать точно. Различные факторы, такие как экономические изменения, социокультурные тенденции, технологические прорывы и политические решения, могут существенно повлиять на потребление электроэнергии, и невозможно учесть все возможные варианты в прогнозах.
2. Сезонные и временные изменения: Электропотребление часто подвержено сезонным и временным колебаниям. Например, летом потребление может возрасти из-за использования кондиционеров, а в праздничные дни или выходные потребление может снизиться. Прогнозирование этих изменений требует учета множества факторов и данных.
3. Развивающиеся технологии: Быстрое развитие новых технологий, таких как электромобили, возобновляемые источники энергии и энергоэффективные устройства, меняет пейзаж потребления электроэнергии. Прогнозирование, как и когда эти технологии будут внедрены, представляет вызов для экспертов.
4. Энергетическая политика и регулирование: Политические решения и изменения в законодательстве в области энергетики могут существенно повлиять на спрос на электроэнергию.
5. Глобальные факторы: Электроэнергия является частью мировой экономики, и множество глобальных факторов, таких как изменения климата, геополитические конфликты и экономические кризисы, могут оказать влияние на ее потребление.

ЗАДАЧА 2 ЭТАП «ПОЧАСОВОЕ ПОТРЕБЛЕНИЕ»

Задача

Разработка сервиса прогнозирования почасового энергопотребления региона на сутки, в МВт*ч

Цель

Разработать надежную и точную модель прогнозирования объема почасового энергопотребления на следующие сутки для Калининградской области с использованием доступных исторических данных и соответствующих переменных.

Описание задачи

В данной задаче необходимо разработать предиктивную модель, которая позволит прогнозировать почасовое энергопотребление региона на следующие сутки на основе имеющихся данных о потреблении электроэнергии в прошлом и соответствующих факторах, влияющих на потребление энергии.

Замечание (!)

Если мы делаем прогноз на сегодня, то у нас есть все данные за вчера и более ранние, но нет данных из будущего. Учесть это при конструировании признаков для модели.

При прогнозировании электропотребления, например, на 2023-04-01 14:00 у нас есть данные фактического электропотребления и фактической погоды/температуры только за вчерашние сутки (т.е. все данные до 2023-03-31 23:59 включительно) и нет данных за 2023-04-01 (кроме прогноза погоды и температуры – они у нас есть на 2023-04-01 до 14:00 включительно)



ДАТАСЕТ

	date	time	target	temp	temp_pred	weather_pred	weather_fact
0	2019-01-01	0	481.510	2.9	2.0	пасм, ветер	ветер
1	2019-01-01	1	462.872	2.9	2.0	пасм, ветер	ветер
2	2019-01-01	2	449.718	2.9	2.0	пасм, ветер	ветер
3	2019-01-01	3	430.908	4.3	2.0	пасм, ветер	ветер, пасм
4	2019-01-01	4	415.163	4.3	2.0	пасм, ветер	ветер, пасм
...

- date** – дата
- time** – время, время представлено в диапазоне 0 – 23, что означает 24 часа в сутках
- target** – фактическое потребление на указанную дату
- temp** – фактическая температура на указанную дату
- temp_pred** – прогноз температуры на указанную дату
- weather_fact** – фактическая погода на указанную дату
- weather_pred** – прогноз погоды на указанную дату

Обучающий датасет

Датасет для обучения. Содержит данные за период

2019-01-01 – 2023-03-31

Публичный тестовый датасет

Представляет продолжение обучающего датасета. Содержит данные за период (и более ранние)

2023-04-01 – 2023-07-31

Приватный тестовый датасет

Содержит данные за период (и более ранние)

2023-08-01 – 2023-09-30

МЕТРИКИ 2 ЭТАП «ПОЧАСОВОЕ ПОТРЕБЛЕНИЕ»

MAE (Главная метрика)

MAPE (Вспомогательная метрика)

R2-score (Вспомогательная метрика)

ОЦЕНКА

Все команды будут ранжированы по метрике MAE, измеренной на приватном тестовом датасете

Что мы от вас ждем

Необходимо предоставить

1. Ноутбук обучения модели с проверкой всех метрик на трейн и тест выборках
2. Модуль `main.py`, который получает на вход путь к приватному тестовому датасету, а на выходе отдает csv-файл с прогнозом в формате **`datetime, predict`** (в файле должны быть только 2 столбца с фиксированными названиями **`[datetime, predict]`** !)

Пример команды: ***`python main.py path_to_private_dataset`***

3. Файл ***`requirements.txt`***, содержащий необходимые библиотеки для работы скрипта. (Не нужно в нем указывать вообще все библиотеки, установленные на вашем компьютере!)
4. Время работы скрипта не более 1 минуты. Размер не более 500 МБ