**Кейс «Моделирование пиковых часов потребления электроэнергии»**

**Час пиковой** **нагрузки** — это час в каждом рабочем дне, в котором потребление электроэнергии в регионе было максимальным.

От того, насколько много субъектов потребляют электроэнергию в каждый час, зависит стоимость МВт.ч, потребленного в этот час.

В случае, если известны часы наибольшего потребления электроэнергии в сутках, субъект (промышленное предприятие) может переносить потребление на другие часы (останавливать производство на некоторое время). Это позволяет оптимизировать расходы на оплату электроэнергии.

Предполагается использование статистики по потреблению электроэнергии по субъектам РФ (Пермский край и Тюменская область) для прогнозирования пиковых часов потребления электроэнергии.

**Перечень источников для использования:**

| **#** | **Ресурс** | **Описание** |
| --- | --- | --- |
| 1 | pik\_chas2022.pdf | Плановые часы пиковой нагрузки на 2021 год для территорий, отнесенных к ценовым зонам оптового рынка  электрической энергии и мощности, и территорий, отнесенных к неценовым зонам оптового рынка  электрической энергии и мощности |
| 2 | Плановое потребление электроэнергии\_Пермский\_край.csv  Плановое потребление электроэнергии\_Тюменская область.csv | Отчет о торгах по субъектам РФ ЕЭС |
| 3 | Фактическое потребление электроэнергии\_Пермский\_край.csv  Фактическое потребление электроэнергии\_Тюменская область.csv | Суммарное фактическое потребление электрической энергии в субъекте РФ |
| 4 | t\_calendar\_days | Трудовой календарь |

**Перечень дополнительных ресурсов для информации:**

| **#** | **Ресурс** | **Описание** |
| --- | --- | --- |
| 1 | https://www.atsenergo.ru/results/market | Отнесение регионов РФ к ценовым / неценовым зонам |
| 2 | https://www.atsenergo.ru/ nreport?rname=fact\_region | Суммарное фактическое потребление электрической энергии в субъекте РФ |
| 3 | https://www.atsenergo.ru/ nreport?rname=trade\_region\_spub | Отчет о торгах по субъектам РФ ЕЭС |

**Цель:** создать модель, позволяющую спрогнозировать пиковые часы для регионов на следующий месяц.

**Результатом вашей работы должно стать решение следующих задач:**

1. Рассчитать часы пиковой нагрузки в регионах
2. Построить эталонную модель, позволяющую относительно простым способом спрогнозировать часы пиковой нагрузки для месяцев в 2022 году
   * Допустимо прогнозирование диапазона из 4-х последовательных часов (т. е. хотя бы один из четырех должен совпасть с фактическим значением)
3. Предложить собственную модель, которая, позволит осуществить ретропрогноз за май-октябрь 2022 в регионе
   * Допустимо прогнозирование диапазона из 4-х последовательных часов (т. е. хотя бы один из четырех должен совпасть с фактическим значением)

По итогам работ необходимо представить отчет, демонстрирующий результаты моделирования и характеристики построенной модели.

***Техническое задание:***

**Описание исходных данных**

Исходные данные содержат в себе 4 таблицы по 2-м регионам (Пермский Край и Тюменская область):

1. Таблица 1 (pik\_chas2022.pdf) – информация о часовых диапазонах для регионов разных ценовых зон (Пермский край и Тюменская область относятся к 1-й ценовой зоне)
2. Таблица 2 (Плановое потребление электроэнергии) – файл, содержащий результаты ежедневных торгов – показатели технологических минимумов производства электроэнергии, плановые объемы производства электроэнергии по станциям (в основном ТЭС), а также плановые объемы потребления электроэнергии
3. Таблица 3 (Фактическое потребление электроэнергии) – Фактическое потребление электроэнергии в субъекте РФ в часовой динамике
4. Таблица 4 (Трудовой календарь) – файл, содержащий отметки о рабочих и праздничных днях в календаре

**Таблица 1 - информация о ценовых диапазонах (описание полей)**

**Описание таблицы 1:**

Данная таблица содержит часовые диапазоны для регионов, входящие в ценовые зоны 1 и 2, а также для отдельных регионов, не входящих ни в какие ценовые зоны.

Пиковым часом может являться только час, входящий в указанные часовые диапазоны соответствующих ценовых зон для каждого месяца.

Например, для Тюменской области в марте пиковыми могут являться только часы, входящие в диапазон с 8-го до 21-го часов.

**Таблица 2 - Плановое потребление электроэнергии**

| **#** | **Поле** | **Описание поля** | **Тип данных** |
| --- | --- | --- | --- |
| 1 | date | Дата | timestamp |
| 2 | P1 | Плановый объем производства (по типам станций), МВт.ч. ТЭС | float |
| 3 | P2 | Суммарные величины технического минимума (по типам станций), МВт.ч. ТЭС | float |
| 4 | P3 | Суммарные величины технологического минимума (по типам станций), МВт.ч. ТЭС | float |
| 5 | P4 | Суммарные величины технического максимума (по типам станций), МВт.ч. ТЭС | float |
| 6 | P5 | Плановый объём потребления, МВт.ч. | float |
| 7 | P6 | Полный плановый объем потребления, МВт.ч. | float |
| 8 | subject | Регион РФ | string |

**Таблица 3 - Фактическое потребление электроэнергии**

| **#** | **Поле** | **Описание поля** | **Тип данных** |
| --- | --- | --- | --- |
| 1 | subject | Регион РФ | string |
| 2 | period | Месяц, в котором производятся измерения | string |
| 3 | date | Дата | timestamp |
| 4 | energy\_consumption | Суммарное фактическое потребление электроэнергии в субъекте РФ, МВт.ч. | float |

**Таблица 4 – Трудовой календарь**

| **#** | **Поле** | **Описание поля** | **Тип данных** |
| --- | --- | --- | --- |
| 1 | id | Идентификатор записи | int |
| 2 | dt | Дата | timestamp |
| 3 | isworkday | Отметка о рабочем дне (Если день рабочий = true) | bool |
| 4 | isholiday | Отметка о праздничном дне (Если день праздничный = true) | bool |

**Часть 1: Предварительная обработка данных**

Необходимо рассчитать пиковые часы, используя данные о фактическом потреблении электроэнергии и интервалы из документа «Плановые часы пиковой нагрузки на 2021 год для территорий, отнесенных к ценовым зонам оптового рынка».

Для этого следует вычислить максимальное значение фактического потребления электроэнергии для каждого дня внутри диапазона, доступного в документе «Плановые часы пиковой нагрузки на 2021 год для территорий, отнесенных к ценовым зонам оптового рынка».

Таким образом, час пиковой нагрузки это:

1. Час с наибольшим значением фактического потребления энергии в течение дня
2. Находящийся в часовом диапазоне для ценовой зоны, соответствующем моделируемому региону

В итоге необходимо получить данные в виде временного ряда вида:

**Регион РФ** -> **Дата** -> **Номер часа пиковой нагрузки**

**Часть 2: Эталонная модель**

Эталонная модель – способ прогнозирования, при котором в качестве прогноза для пиковых часов используются наиболее часто встречающиеся значения соответствующего месяца прошлого года, объединенные в 4-часовой диапазон.

Для построения эталонной модели необходимо:

1. Вычислить статистику по пиковым часам за предыдущий год (доля значений пиковых часов за весь месяц, соответствующих каждому часу с 0-го по 23-й)
2. Определить 4-часовой диапазон для месяца, для которого сумма долей максимальна
3. Если пиковое значение попадает на единственный час, то для диапазона необходимо брать значения, предшествующие ему.
4. Использовать данный диапазон для прогнозирования значений за соответствующие месяцы 2022 г.

**Таблица 4 – Статистика по пиковым часам региона в 2021 г. (пример)**

|  | **март** | **апрель** | **май** | **июнь** |
| --- | --- | --- | --- | --- |
| **0** | 0% | 0% | 0% | 0% |
| **1** | 0% | 0% | 0% | 0% |
| **2** | 0% | 0% | 0% | 0% |
| **3** | 0% | 0% | 0% | 0% |
| **4** | 0% | 0% | 0% | 0% |
| **5** | 0% | 0% | 0% | 0% |
| **6** | 0% | 0% | 0% | 0% |
| **7** | ***10%*** | 0% | 0% | 0% |
| **8** | ***17%*** | 0% | 0% | ***70%*** |
| **9** | 0% | ***15%*** | 0% | ***5%*** |
| **10** | ***25%*** | ***25%*** | 0% | 0% |
| **11** | ***15%*** | 0% | 0% | 0% |
| **12** | ***5%*** | 0% | 0% | 0% |
| **13** | 0% | 0% | 0% | 0% |
| **14** | 0% | 0% | ***94%*** | 0% |
| **15** | 0% | ***14%*** | 0% | 0% |
| **16** | 0% | 0% | 0% | 0% |
| **17** | 0% | 0% | 0% | 0% |
| **18** | ***23%*** | 0% | 0% | 0% |
| **19** | 0% | 0% | 0% | 0% |
| **20** | 0% | ***30%*** | 0% | ***10%*** |
| **21** | ***5%*** | ***16%*** | ***6%*** | ***15%*** |
| **22** | 0% | 0% | 0% | 0% |
| **23** | 0% | 0% | 0% | 0% |

Согласно данной таблице прогнозом для марта 2022 года будет диапазон от 8-го до 10-го часов (так как в сумме это 57% всех значений марта 2021).

Прогноз для апреля – часы с 18-го до 21-го (46% всех значений апреля 2021).

Прогноз для мая – часы с 11-го до 14-го (94% всех значений мая 2021).

Прогноз для июня – часы с 6-го до 9-го (75% всех значений июня 2021).

Тот прогноз, который выдает эталонная модель – это константа, неизменная в течение всего месяца.

**Часть 3: Моделирование и прогнозирование часов пиковой нагрузки**

Необходимо предложить модель, способную спрогнозировать час пиковой нагрузки, а также диапазон из 4-х последовательно расположенных часов.

Прогноз необходимо построить в дневной динамике на месяц вперед.

Для оценки качества построенной модели предлагается осуществить ретропрогноз на май, июнь, июль, август, сентябрь и октябрь 2022 года.

Для этого необходимо рассчитать прогноз для рассматриваемого месяца оценив модель на данных до него.

**Прогноз считается точным**, если фактическое значение часа пиковой нагрузки совпало с одним из четырех подряд идущих часов из прогнозируемого диапазона (прогнозируемый час пиковой нагрузки + 3 соседних часа).

*Предлагается несколько способов моделирования часов пиковой нагрузки:*

1. Моделирование и прогнозирование пиковых часов напрямую через модели машинного обучения
2. Моделирование и прогнозирование классов (утро, день, вечер и т. д.)
3. Моделирование и прогнозирование фактического потребления электроэнергии и вычисление на основе этого показателя часов пиковой нагрузки
4. Другие способы

**Часть 4: Отчет о моделировании часов пиковой нагрузки**

По результатам выполнения задач из частей 1 - 3 предлагается написать отчет, в котором необходимо описать особенности построенных моделей и сравнить их качество с качеством эталонной модели.

Качество модели для каждого месяца (Qi) вычисляется как:

,

где xi – число рабочих дней в i-м месяце, для которых хотя бы 1 из 4-х спрогнозированных часов совпал с фактическим пиковым часом,

Yi – общее число рабочих дней в i-м месяце.