## Задача 3. Прогнозирование потребления тепла Dự báo tiêu thụ nhiệt

В уездном городке У есть проблема: в конце каждой осени приходит зима. А с зимой приходят холода. А когда на улице холодно, жители просят, чтобы дома было тепло. При этом каждый дом потребляет различное количество теплоты, и хотелось бы спрогнозировать, какое же будет потребление по дому в предстоящую зиму. Ở thị trấn Wu có một vấn đề: cứ cuối thu đông lại đến. Và cùng với mùa đông là thời tiết lạnh giá. Và khi bên ngoài trời lạnh, người dân mong muốn được sưởi ấm trong nhà. Đồng thời, mỗi ngôi nhà tiêu thụ một lượng nhiệt khác nhau và tôi muốn dự đoán mức tiêu thụ của hộ gia đình trong mùa đông tới.

В большинстве домов городка установлены счётчики теплоты, которые отдают в автоматическом режиме данные каждый день в течение всего отопительного сезона. Hầu hết các ngôi nhà trong thị trấn đều được lắp đặt đồng hồ đo nhiệt độ, tự động cung cấp dữ liệu hàng ngày trong suốt mùa sưởi ấm.

Необходимо спрогнозировать суммарное потребление теплоты каждым домом на следующий отопительный сезон, чтобы городской поставщик отопления был уверен, что ресурсов у него хватит. Cần dự đoán tổng lượng nhiệt tiêu thụ của từng ngôi nhà trong mùa sưởi tiếp theo để nhà cung cấp hệ thống sưởi của thành phố chắc chắn rằng họ có đủ nguồn lực.

Данные для построения модели:

1. Архив температуры наружного воздуха (файл «*температура.csv*») со следующими параметрами: Lưu trữ nhiệt độ không khí bên ngoài (file “nhiệt độ.csv”) với các thông số sau:
   * *date\_start* - дата и время начала периода измерений; ngày và giờ bắt đầu khoảng thời gian đo;
   * *date\_end* - дата и время окончания периода измерений; ngày và giờ kết thúc khoảng thời gian đo;
   * *temp* - средняя температура за период; nhiệt độ trung bình trong khoảng thời gian;
   * *temp\_max* - наибольшая температура за период; nhiệt độ cao nhất trong khoảng thời gian;
   * *temp\_min* - наименьшая температура за период; nhiệt độ thấp nhất trong khoảng thời gian đó;
   * *humidity* — влажность. độ ẩm.
2. Характеристики домов (файл «*характеристики\_домов.csv*») со следующими параметрами: Đặc điểm nhà (file “house\_characteristics.csv”) với các thông số sau:
   * *address\_uuid* - уникальный идентификатор адреса; mã định danh địa chỉ duy nhất;
   * *build\_year* - год постройки; năm xây dựng;
   * *floor\_number* - количество этажей; số tầng;
   * *residential\_area* - общая жилая площадь; tổng diện tích sinh hoạt;
   * *wall\_type* - тип наружных стен; loại tường bên ngoài;
   * *roof\_area\_total* - общая площадь крыши tổng diện tích mái nhà
   * *roof\_area\_metal* - площадь металлической кровли ; diện tích của mái kim loại;
   * *roof\_area\_web* - площадь кровли из рулонных материалов; diện tích mái làm bằng vật liệu cuộn;
   * *roof\_area\_piece\_goods* - площадь кровли из штучных материалов. diện tích mái được làm bằng vật liệu tấm.
3. Суточных архив показаний тепловычислителей (файл «*объём\_теплоты.csv*») со следующими параметрами; Lưu trữ số liệu đo nhiệt kế hàng ngày (tệp “heat\_volume.csv”) với các thông số sau;
   * *address\_uuid* - уникальный идентификатор адреса; mã định danh địa chỉ duy nhất;
   * *date* - дата, за которую снимается показание; ngày thực hiện phép đo;
   * *value* - значение потреблённой домом теплоты за указанную дату; giá trị nhiệt lượng mà ngôi nhà tiêu thụ trong ngày được chỉ định;
   * *is\_unreliable* - флаг достоверности показаний. cờ về độ tin cậy của kết quả đo.

Исходные данные:

1. Файл с «прогнозом» температуры на следующий отопительный сезон ***/mnt/source/temperature.csv*** с параметрами, аналогичными файлу «*температура.csv*» из данных для построения модели. Tệp có “dự báo” nhiệt độ cho mùa nóng tiếp theo /mnt/source/template.csv với các tham số tương tự như tệp “nhiệt độ.csv” từ dữ liệu để xây dựng mô hình

Выходные данные:

1. Файл с суммой потреблённого тепла по каждому дому ***/mnt/result/volume.csv*** со следующими параметрами: File lượng nhiệt tiêu thụ của mỗi ngôi nhà /mnt/result/volume.csv với thông số sau:
   * *address\_uuid* - уникальный идентификатор адреса; mã định danh địa chỉ duy nhất;
   * *volume* - сумма спрогнозированных показаний по дому по датам из исходных данных на основании указанного прогноза погоды. tổng số chỉ số dự đoán cho ngôi nhà theo ngày từ dữ liệu nguồn dựa trên dự báo thời tiết đã chỉ định.

Примечания:

* Формат всех файлов CSV: с заголовком, разделители - «;».
* Считается, что показания «прогноза» температуры относятся к отопительному сезону, следующему за данными для построения модели. Các số liệu “dự báo” nhiệt độ được coi là đề cập đến mùa nóng theo dữ liệu xây dựng mô hình.
* Дополнительные данные о домах даны, чтобы можно было кластеризовать временные ряды показаний потребления тепла по характеристикам домов. Dữ liệu bổ sung về ngôi nhà được cung cấp để cho phép phân nhóm các chỉ số tiêu thụ nhiệt theo chuỗi thời gian theo đặc điểm của ngôi nhà.

На проверку отправляется архив, в котором присутствует:

* Файл ***README.pdf*** с кратким описанием применённого решения.
* Исходные тексты программы, формирующей модель, в папке ***/src-train***.
* Исходные тексты программы, преобразующей исходные данные в выходные на основании заранее сформированной (обученной) модели в папке ***/src-predict***.
* Собственно модель в папке ***/model***.
* Файл ***Dockerfile***, в котором создаётся контейнер и запускается программа для преобразования исходных данных в выходные (из */src-predict*).

Требование самостоятельно описывать окружение через Dockerfile необходимо для унификации проверки. Пример файла «*docker-compose.yml*» для запуска задания на проверку: Yêu cầu mô tả độc lập môi trường thông qua Dockerfile là cần thiết để thống nhất xác minh. Một ví dụ về tệp “docker-compose.yml” để chạy công việc xác minh:

|  |
| --- |
| version: "3.1"  services:  app:  build: .  volumes:  - .:/app  - ${PATH\_SOURCE}:/mnt/source  - ${PATH\_RESULT}:/mnt/result |

, где ${PATH\_SOURCE} - путь до каталога с исходными данными, ${PATH\_RESULT} - путь до каталога с выходными данными для проверки.

То есть запуск программы будет осуществляться со следующими характеристиками: Tức là chương trình sẽ được khởi chạy với các đặc điểm sau:

* ОС - Linux, настроенная через Dockerfile для запуска программы;
* Данные их архива на проверку примонтированы в */app*;
* Соответственно, модель находится по пути */app/model*;
* Исходные данные примонтированы в */mnt/source*;
* Каталог для сбора выходных данных примонтирован в */mnt/result*.

Запуск контейнера будет осуществляться через команду «docker compose up --build» с заданными переменными окружения PATH\_SOURCE и PATH\_RESULT.

**Если у Вас возникают трудности с написанием Dockerfile, опишите более подробно, как запустить программу predict для формирования выходного файла. Nếu bạn gặp khó khăn khi viết Dockerfile, vui lòng mô tả chi tiết hơn cách chạy dự đoán để tạo tệp đầu ra.**