

Данные и задания на Марафон

1. КНВШ

Сведения об основных и дополнительных образовательных программах, реализуемых организациями, расположенными на территории Санкт-Петербурга

Данные:

– Информация о мероприятиях, проводимых научными и образовательными организациями Санкт-Петербурга, JSON: <https://researchinspb.ru/api/v1/public/event/>

– Информация о вакансиях научных и образовательных организаций Санкт-Петербурга, JSON: <https://researchinspb.ru/api/v1/public/vacancy/>

2. ИАЦ

Сопоставление баз и интеллектуальный поиск адресов

Данные:

– api.petersburg.ru

– <https://portal.iac.spb.ru/~pJ4Kq>

– <https://portal.iac.spb.ru/~5MXZd>

Задание:

– Создать алгоритм на базе ИИ, который получает на вход два списка адресов с уникальными ID. Адреса в списках идентичны по смыслу, но различаются по написанию (например, "Невский проспект" и "Невский пр"). Алгоритм должен создать таблицу соответствия ID между этими списками.

– Разработать веб-приложение (фронтенд + бэкенд), реализующее нечёткий поиск по адресам. Система должна находить и предлагать варианты адресов при неполном вводе (например, "невс про 15") или при наличии опечаток.

Ожидаемый результат: Рабочий прототип с открытым API для интеграции, демонстрирующий оба компонента задачи.

3. Stream Telecom

Набор обезличенных данных об активности пользователей при переходе по ссылкам.

Данные:

– <https://disk.yandex.ru/d/Wz3VhmyemvSbZQ>

Задание:

На базе полученного набора обезличенных данных об активности пользователей при переходе по ссылкам необходимо реализовать Dashboard, который обеспечит расчет и визуализацию следующих показателей:

– Активность в течение 4 часов с момента создания кампании.

–Динамика кликов по дням и месяцам (абсолютное количество и % соотношение от общего количества).

–Динамика создания кампаний (по дням в разрезе недели и по месяцам в разрезе года).

–Скорость реакции клиентов (общее).

–Географическое распределение клиентов (тепловая карта).

–Горячие точки клиентов (географические зоны с наибольшей активностью).

–Время активности клиентов в течение дня с учетом часового пояса.

–Прогнозируемое лучшее время для отправок.

Описание набора данных

Набор данных представляет из себя три csv-файла:

–**regions** – список регионов и их идентификаторов:

region_id – идентификатор региона.

name – название региона.

–**campaign** – список обезличенных рекламных кампаний, в рамках которых пользователи осуществляли переходы по ссылкам:

id – идентификатор кампании.

created_at – дата и время создания кампании.

–**clicks** – список переходов по ссылкам пользователей:

click_date – дата перехода по ссылке.

click_time - время перехода по ссылке.

uid – идентификатор перехода.

member_id – обезличенный идентификатор пользователя.

campaign_id – идентификатор кампании в рамках которой пользователь перешел по ссылке.

region – идентификатор региона пользователя.

OS – информация об операционной системе устройства пользователя с которого был осуществлен переход.

browser – информация о браузере пользователя.

user_agent - user_agent пользователя.

language – локаль браузера пользователя.

device – устройство пользователя.

4. СПбГЭТУ «ЛЭТИ»

В рамках Марафона СПбГЭТУ «ЛЭТИ» предоставляет данные приемной кампании, расписание занятий, учебные планы с перечнем дисциплин и информацией о вакансиях «Работа в России».

Данные:

–<https://digital.etu.ru/api/docs-public/>

– <https://trudvsem.ru/opendata/datasets>

5. Омега

Информация из медицинской информационной системы/ Предоставленная база данных МИС содержит обезличенные данные о 144000 пациентах, 1100 медицинских специалистах и более 3 млн. записей о посещениях пациентов медицинской организации с диагнозами и сведениями о заболеваниях пациентов.

Данные:

– URL: https://1cbase.gkomega.ru/digitalspb/hs/med_api/patients_list Описание: Возвращает список пациентов медицинской организации

– URL: https://1cbase.gkomega.ru/digitalspb/hs/med_api/specialist_list Описание: Возвращает список специалистов медицинской организации.

– URL: https://1cbase.gkomega.ru/digitalspb/hs/med_api/patient_receptions Описание: Возвращает список посещений пациентов с информацией о дате, диагнозе и специалисте.

– URL: https://1cbase.gkomega.ru/digitalspb/hs/med_api/ping Описание: Проверка доступности API. Возвращает HTTP-статус 200 в случае успешного вызова.

Задание:

– Экстренная медицинская помощь

Платформа для служб скорой помощи, которая оперативно получает историю посещений пациентов (/patient_receptions), чтобы быстрее поставить предварительный диагноз. Это может ускорить оказание помощи в критических ситуациях.

– Аналитика здоровья города

Инструмент для городских властей, использующий агрегированные данные о посещениях врачей (/patient_receptions) для выявления всплеск заболеваний, отслеживания сезонных эпидемий и оценки работы медицинских учреждений.

– Медицинский маршрут

Сервис, который помогает пациентам записываться к нужным специалистам, анализируя их диагнозы и предыдущие посещения врачей.

Использует /patient_receptions для проверки истории пациента и /specialist_list для поиска подходящих врачей.

6. Лартех

Данные:

– <https://disk.yandex.ru/d/ek2MTy6yOVr2HA>

Задание:

– Создать серверное ПО для анализа и визуализации истории показаний приборов учета электроэнергии. Анализ заключается в кластеризации, то есть разбиении потребителей электроэнергии на группы (кластеры) по определённому признаку и в выявлении аномалий.

Признаками для кластеризации могут быть, например, профиль потребления, динамика потребления, тип электрической нагрузки.

Входными данными для ПО будут являться XLS-файлы с получасовыми и суточными показаниями массива умных электросчётчиков. Каждый счётчик принадлежит уникальному потребителю электроэнергии.

Ожидаемые результаты:

1. Создание и описание методов и признаков кластеризации потребителей электроэнергии;
2. Список аномалий, то есть счётчиков, не попавших ни в один кластер, с описанием аномалии;
3. Выбор способа визуализации истории показаний массива счётчиков, их кластеров и его объяснение (почему именно такой способ наиболее информативен).

«Умные электросчётчики». Генерируют и передают «большие данные» следующего вида:

1.Получасовые показания - суммарные по всем трём тарифам измерения активной (A) и реактивной (P) энергии в прямом (+) и обратном (-) направлениях один раз в полчаса;

2.Суточные показания – измерения A+, A-, P+ и P- раздельно по трём тарифам (T0, T1 и T2) один раз в сутки.

Получасовые и суточные данные доступны в виде XLS-файлов с указанием даты, времени (для получасовых данных), серийного номера счётчика и значений вышеперечисленных показаний.

Каждый счётчик принадлежит уникальному потребителю электроэнергии.



–Создать серверное ПО для анализа и визуализации истории показаний приборов учета электроэнергии. Анализ заключается в поиске потребителей-нарушителей, занимающихся в жилых домах (квартирах) майнингом криптовалюты и (или) имеющих ферму из серверов.

Входными данными для ПО будут являться XLS-файлы с получасовыми и суточными показаниями массива умных электросчётчиков. Каждый счётчик принадлежит уникальному потребителю электроэнергии.

Ожидаемые результаты:

1. Создание и описание алгоритмов поиска нарушителей указанного типа (майнеры/серверы);
2. Список счётчиков из предоставленных данных, которые могут относиться к нарушителям указанного типа;
3. Выбор способа визуализации истории показаний массива счётчиков с удобным для оператора графическим отображением нарушителей.