

цифровой
прорыв 

сезон: III

КЕЙС

Правительство Санкт-Петербурга
Комитет по информатизации и связи



Разработка модели по
устранению и
изменению
некорректных адресов
из городских баз данных



Министерство
экономического развития
Российской Федерации



Кейсодержатель

Правительство Санкт-Петербурга
Комитет по информатизации и связи

01 Сфера деятельности

Комитет проводит государственную политику Санкт-Петербурга в сфере цифровой трансформации, информации, информационных технологий и защиты информации, связи, а также координирует деятельность исполнительных органов государственной власти Санкт-Петербурга в указанной сфере.

02 Краткое описание кейса

Разработка модели по точному определению адреса из городских баз данных с применением методов обработки естественного языка (NLP).



Сайт организации

<https://kis.gov.spb.ru/>

Постановка задачи

Участникам хакатона при помощи методов искусственного интеллекта и данных, включающих полные адреса домов и зданий г. Санкт-Петербурга, предстоит построить модель, которая сможет адаптироваться к задачам определения корректного адреса, содержащегося в исходной базе данных.

Ограничения: не использовать сторонние API для определения адреса. Не использовать в обучении модели координаты объекта.

Модель должна уметь до обучаться/переобучаться, при внесении изменений в эталонные данные (соответствие критериям: Адаптивность/Масштабируемость)



Министерство
экономического развития
Российской Федерации

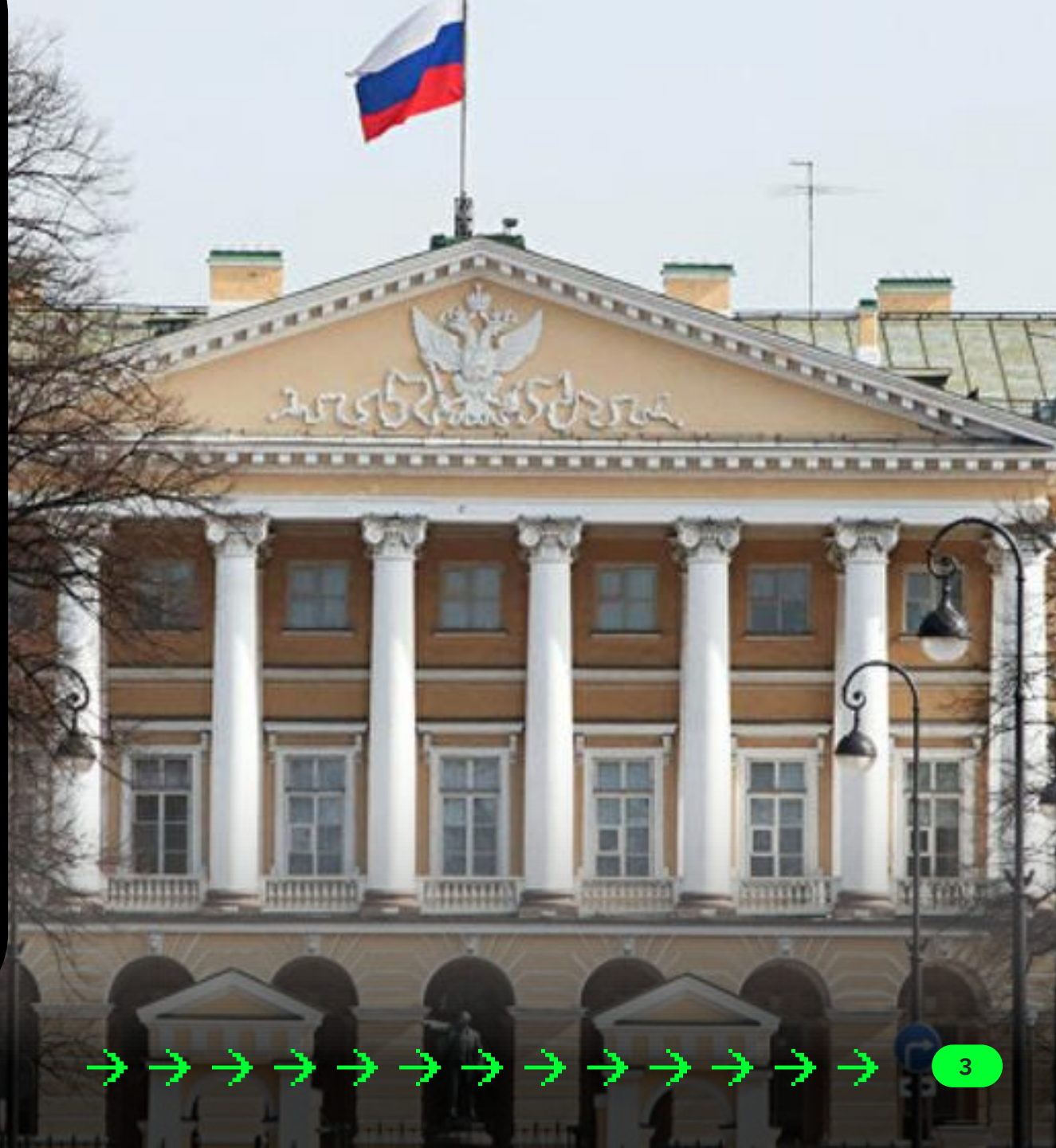


цифровой
прорыв

сезон: III



3



Проблематика

Особенностью задачи по определению адреса является работа с наборами данных с адресами разных форматов и произвольно введенными адресами, что усложняет процесс построения качественных моделей.

Например: “аптерский 18 спб”, “Санкт-Петербург, Яхтеная у. 18-16-4”, “Петергоф Ломоносовский 3 корп.1”

Проблема в неправильном определении адресов и координат напрямую влияет на создание качественных датасетов и функциональность сервисов.

Решение

Результат должен быть представлен в виде работающего API и в виде заполненного датасета.

Минимальное кол-во введенных символов в Query должно быть не менее 4-х, независимо от регистра.

Пример запроса: «аптерский 18 спб»

Response должен содержать все определенные моделью адреса, но не более 10 в порядке, от наиболее точного определения к менее точному, представленных в виде массива, а при максимальной точности результата, не более 3-х ответов. При отсутствии результата, должен быть ответ, что адрес не найден. Пример правильного ответа (адрес найден):

```
{
  "success": true,
  "query": [
    "address": "аптерский 18 спб"
  ],
  "result": [
    {
      "target_building_id": 209676,
      "target_address": "г.Санкт-Петербург, Аптекарский проспект, дом 18, литера А",
      "target_coordinates": [59.974953, 30.31938]
    }
  ]
}
```

Приложение упаковано в Docker и после разворачивания готово к работе.

Разработанная участниками модель, поможет решить вопрос совершенствования методов определения корректного адреса из Государственных информационных систем, позволит объединить большое количество разобщенных городских баз данных, что является важной задачей как для проектов «Комитета по информатизации и связи», ГУП СПб «Информационно-аналитический центр», так и для других разработчиков.

Наличие инструментов для точного определения адресов, является важным и необходимым аспектом, как для создания приложений и сервисов на базе Государственных информационных систем, так и самих Государственных информационных систем.

Стек технологий, обязательных к использованию

Необходимые данные, дополнения, пояснения, уточнения

01

Python, FastApi, Docker

02

Для выполнения задания рекомендуем изучить
<https://fias.nalog.ru/> - ФЕДЕРАЛЬНАЯ
ИНФОРМАЦИОННАЯ АДРЕСНАЯ СИСТЕМА



Оценка

→ Для оценки решений применяется метод экспертных оценок и автоматизированные средства оценивания.

→ Жюри состоит из отраслевых экспертов и/или представителей кейсодержателя.

→ На основании описанных ниже характеристик, жюри выставляет оценки 0-3 балла.

→ Итоговая оценка определяется как сумма баллов всех экспертов: технического, отраслевого и/или представителя кейсодержателя, как значение, выданное автоматизированными средствами оценивания, либо как итоговый балл жюри, умноженный на оценку автоматизированной системы.



Отраслевой эксперт и/или представитель кейсодержателя оценивает решение по следующим критериям:

01

Релевантность поставленной задаче
(команда погрузилась в отрасль,
проблематику; предложенное решение
соответствует поставленной задаче;
проблема и решение структурированы)

02

Уровень
реализации
(концепция/
прототип и т.д.)

03

Проработка
пользовательских
историй (UX/ UI)

04

Реализация в решении
требований Заказчика

05

Выступление команды (умение
презентовать результаты своей работы,
строить логичный, понятный и интересный
рассказ для презентации результатов своей
работы)



Технический эксперт оценивает решение по следующим критериям:

01

Запускаемость кода

02

Обоснованность выбранного метода (описание подходов к решению, их обоснование и релевантность задаче)

03

Точность работы алгоритма (возможность оценить формальной метрикой с обоснованием выбора)

04

Адаптивность/
Масштабируемость

05

Отсутствие в решении импортного ПО и библиотек, кроме свободно распространяемого с обоснованием выбора

06

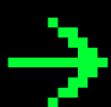
Наличие интеграционных интерфейсов, в первую очередь интерфейсов загрузки данных

Автоматизированные средства оценивания точности работы предложенных участниками алгоритмов (решений) выставляют оценку в диапазоне 0-1, где 1 равно 100% точности работы решения.

Итоговая оценка определяется как итоговый балл жюри, умноженный на оценку автоматизированной системы.

цифровой прорыв

сезон: III



Министерство
экономического развития
Российской Федерации

