

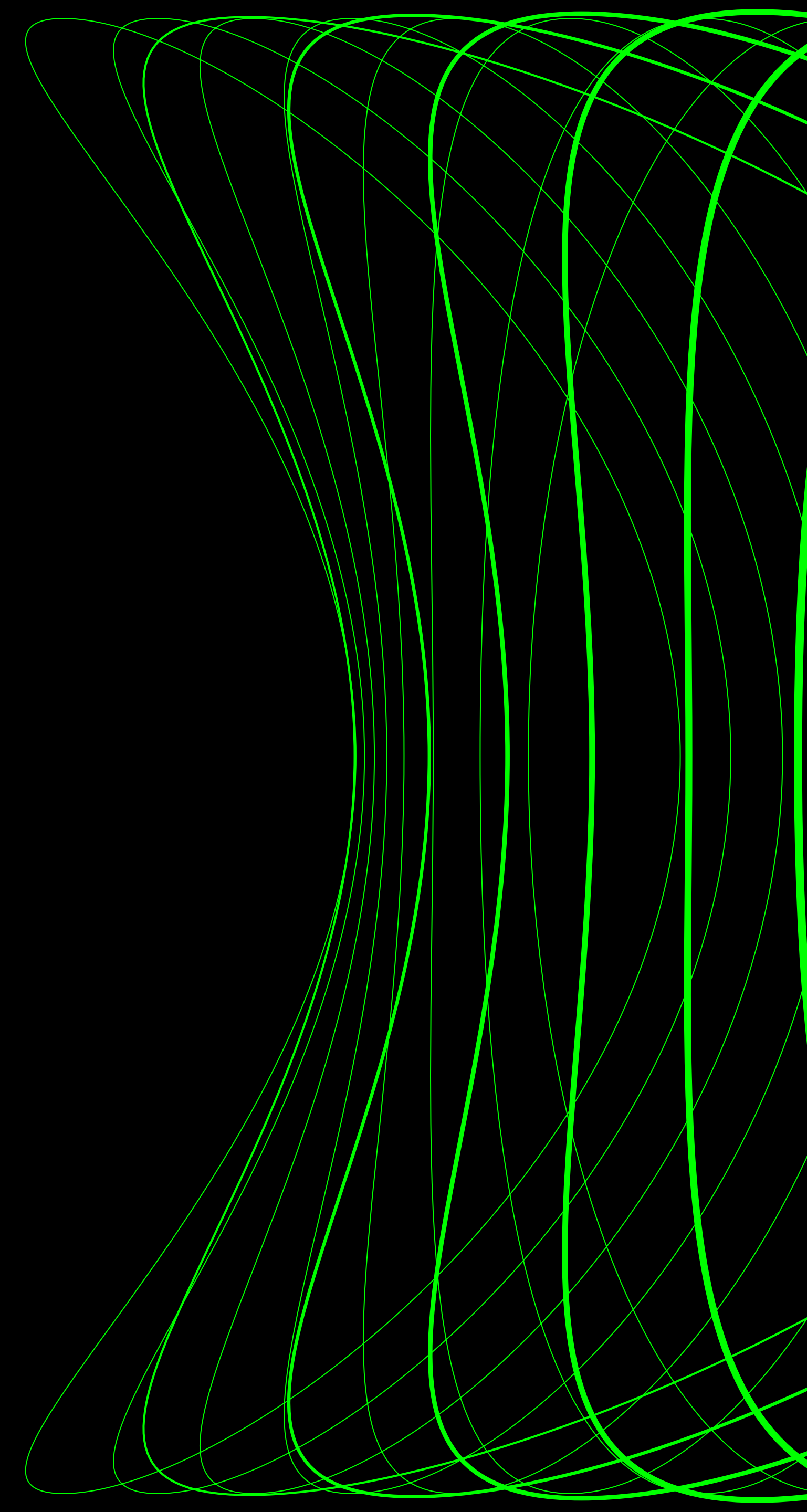
цифровой
прорыв

сезон: или

КЕЙС

ООО «ВК» (VK. COMPANY)

Прогноз, осуществления целевого действия
пользователем в зависимости от его кликстрима, считая
метрикой качества построенной модели ROC-AUC score



Кейсодержатель

Название организации

ООО «ВК»

Сфера деятельности

VK развивает экосистему сервисов, которые помогают миллионам людей решать повседневные задачи онлайн

Краткое описание кейса

Предсказать, будет ли осуществлено целевое действие пользователем в зависимости от его кликстрима, считая метрикой качества построенной модели ROC-AUC score

Сайт организации

<https://vk.company/ru/>

Постановка задачи

Классическая задача бинарной классификации на структурированных табличных и неструктурированных текстовых источниках данных. Каждый объект обучающей выборки характеризуется набором интернет-сессий (непрерывные последовательности посещенных пользователем сайтов).

Проблематика

Решение задачи осложняется низкой долей положительного класса, наличием пропусков, шумом в признаках и целевой переменной, необходимостью экстракции признаков из неструктурированного текстового источника.

Решение

Решение кейса представляет собой прототип системы которая определяет, будет ли совершено целевое действие или нет. При оценке будет учитываться не только объективное качество модели, но и объяснимость критериев..

Необходимые данные, дополнения /пояснения/уточнения

Датасет содержит 3 колонки:

1) ID объекта

2) Список url-ов в формате «url counter», где counter – счётчик, показывающий, сколько раз встретился данный url в кликстриме человека

3) Список title-ов в формате «title counter», где counter – счётчик, показывающий, сколько раз встретился данный

title в кликстриме человека

Требования к результату и критерии оценивания

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, отраслевых и/или представителей Кейсодержателя, умноженная на значение метрики автоматического средства оценивания.



Технический эксперт оценивает решение по следующим критериям:

- 1. Запускаемость кода**
- 2. Обоснованность выбранного метода** (описание подходов к решению, их обоснование и релевантность задаче)
- 3. Точность работы алгоритма** (возможность оценить формальной метрикой с обоснованием выбора)
- 4. Адаптивность/
Масштабируемость**
- 5. Отсутствие в решении импортного ПО и библиотек, кроме свободно распространяемого с обоснованием выбора**

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.

Ссылки на датасет



https://ai-data.obs.ru-moscow-1.hc.sbercloud.ru/ии_по_следам_пользователей.zip

**цифровой
прорыв** 

сезон: III

