DATA FUSION CONTEST 2023

Meetup #1

16.02.2023

Welcome!

1. Про соревнование
   - Обзор
   - Таймлайны
   - Механика
   - Bonus
2. Про подходы и бейзлайны
3. Q&A

Track на **ODS.AI**:
Data Fusion Contest 2023

DATA FUSION CONTEST 2023

Часть 1 — обзор

Тема соревнований — атаки на ML модели

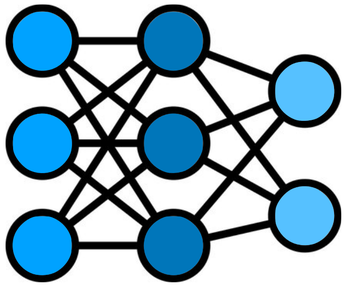
Что значит “атаки”?

1. Классификация ~ as usual

Входные данные

Подаем на вход модели

ML модель



Получаем вероятности

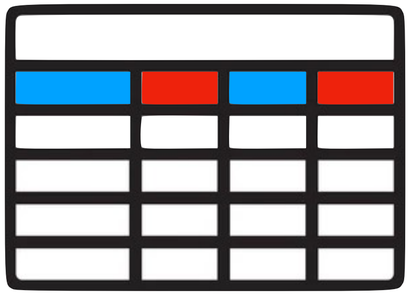
Результат модели

V **P = 0.92**

X P = 0.08

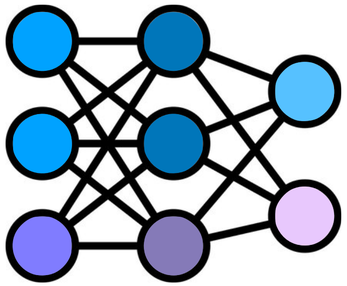
2. Атакуем классификатор через данные

Входные данные



Портим входные данные

ML модель



Меняем вероятности

Результат модели

V P = 0.11

X **P = 0.89**

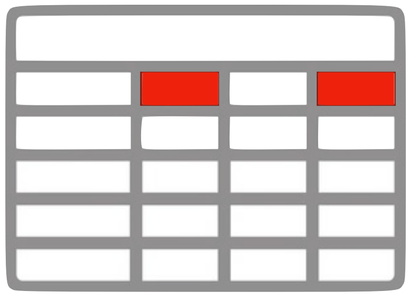
Common ~~sense~~ ground

Для участников всё должно быть привычно 🤗

1. Данные соревнования

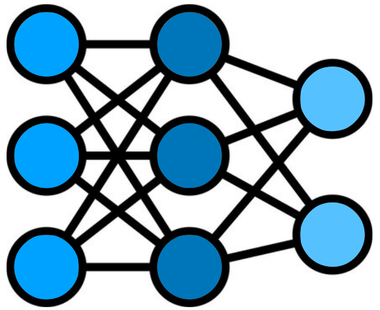
+

2. Механизм атаки



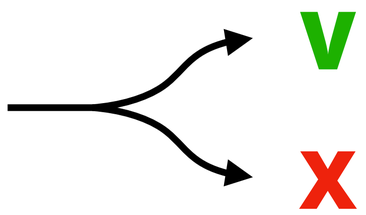
+

3. Модель для атаки



+

4. Целевая переменная

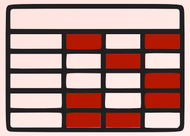


Чтобы все получилось — нужно зафиксировать основу соревнования 👀

Как все устроено?

Задача 1, “Атака”

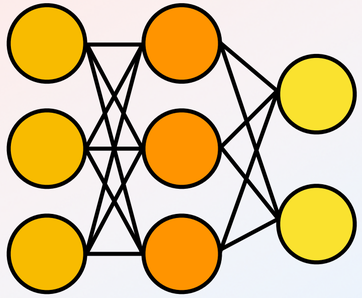
- Берём данные и модель
- Учимся вносить правки в данные
- Есть ограничения на сами правки

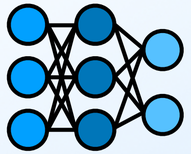


Отправляем свои версии сгенерированных атак на модель организаторов

табличные .csv файлы

ML модель



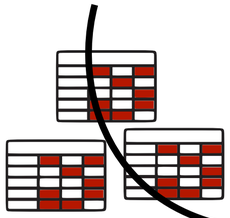


Отправляем свои модели, защищаясь от заготовленных атак организаторов

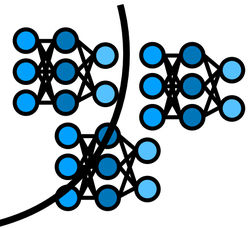
.zip архивы с моделями

Задача 2, “Защита”

- Берем данные и/или модель
- Делаем новую защищенную модель
- Стараемся не просесть по качеству



Турниры



- Турниров 2: промежуточный и финальный
- Все призовые — за победу в турнирах
- На турнирах те же задачи, но на новом наборе данных (формат и размер — те же)

- Отбираем топ-10 команд с каждой задачи
- Запускаем их решения друг против друга
- В каждой задаче появляется по новому лидерборду (с теми же метриками)

DATA FUSION CONTEST 2023

Часть 2 — таймлайны

Timeline соревнования

У каждой задачи будет 5 лидербордов

Public, фаза 1	Public, фаза 2	Турнир 1	Private, фаза 2	Турнир 2

Последовательность открытия лидербордов

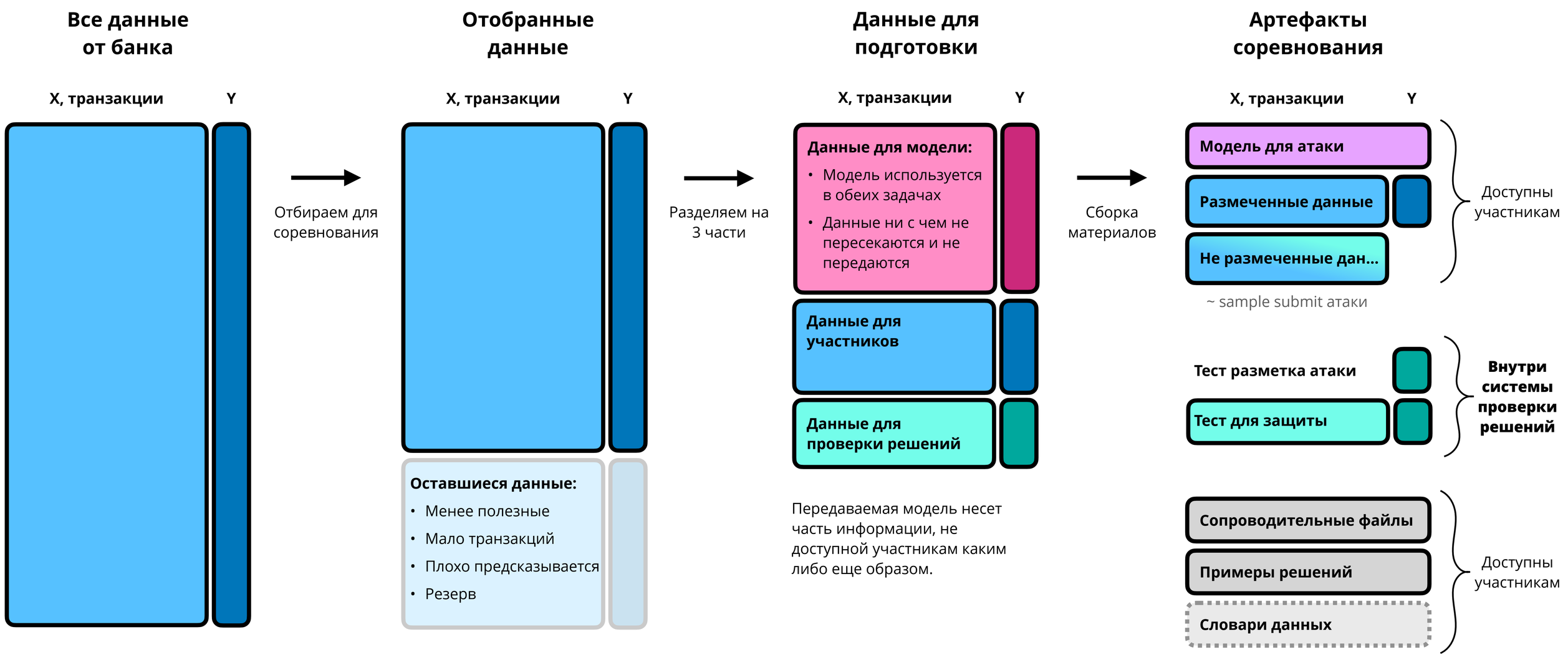
Зеленые — те, где идет активность
Бирюзовые — с замороженным (фиксированным) лб
Пустые — еще не существующие лб.

	Public, фаза 1	Public, фаза 2	Турнир 1	Private, фаза 2	Турнир 2
с 07.02 00:00 по 05.03 23:59	Открыт, основной лидерборд	Недоступен	Недоступен	Недоступен	Недоступен
с 06.03 00:00 по 11.03 19:59	Открыт, статичен (зафиксирован)	Открыт, основной лидерборд	Недоступен	Недоступен	Недоступен
с 11.03 20:00 по 02.04 23:59	Открыт, статичен (зафиксирован)	Открыт, основной лидерборд	Открыт, статичен (зафиксирован)	Недоступен	Недоступен
с 03.04 12:00 по 14.04 19:59	Открыт, статичен (зафиксирован)	Открыт, статичен (зафиксирован)	Открыт, статичен (зафиксирован)	Открыт, статичен (зафиксирован), основной	Недоступен
с 14.04 20:00	Открыт, статичен (зафиксирован)	Открыт, статичен (зафиксирован)	Открыт, статичен (зафиксирован)	Открыт, статичен (зафиксирован), основной	Открыт, статичен (зафиксирован)

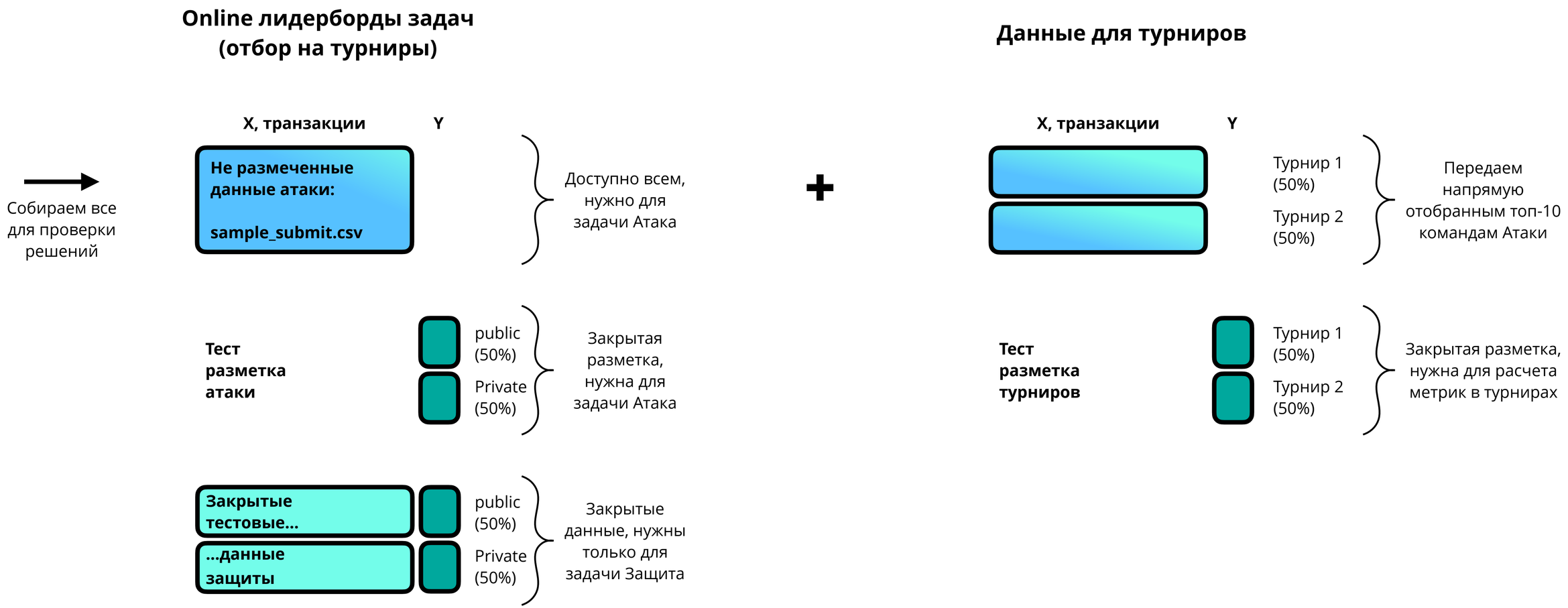
DATA FUSION CONTEST 2023

Часть 3 — механика

Круговорот данных соревнования



Круговорот тестовых данных



Как работают турниры?

Как работают турниры?

1. Квалификация

- 🌑 Берём топ-10 по лидербордам обеих задач
- 👋 Передаём атакующим турнирный sample
- ⏰ За 24 часа получаем от атакующих сабмиты на турнир
  если не прислали — не участвуют в турнире
- 🍿Защищающиеся спокойно ждут начала
  им нужно было только выбрать свои “финальные” решения

Участвуют 20 команд: **10 vs 10**
Запускаются друг против друга 40 решений: **20 vs 20**

Как работают турниры?

2. Запуск

- 💪 Каждое из 20 решений атаки запускается против каждого из 20 решений защиты
- 👷 Каждый запуск — аналогичен запуску решения защиты:

  1. есть исходные транзакции и метки к ним
  2. есть атакованные транзакции от атакующих
  3. есть модель защищающихся, запускающаяся на (1.) и (2.)

- 🧮 Для каждой пары решений сохраняем их ROC-AUC:

  — на исходных данных
  — на атакованных данных

Как работают турниры?

3. Подсчет результатов

- 🤷‍♀️ Каждой из пар запусков считаем их “родные” метрики:

  — **ROC-AUC diff** для атакующих

  — **Mean Harm ROC-AUC** для защищающихся

- ⚙️ Каждому из 40 решений считаем его сводный результат:
  — обычное **среднее арифметическое** по его “родной метрике”

- 📊 Считаем итоговый результат каждой из команд каждой задачи:
  — лучший результат из их двух решений

… уже можно было бы публиковать 👀

Как работают турниры?

4. Оглашение результатов

- 🎉 Оглашаем результаты на ивенте

  — митап 11 марта для первого турнира

  — конференция Data Fusion 2023 для финального турнира

- 💡 Публикуем лидерборд сразу после оглашения результатов

  — отдельный лидерборд с 10 атакующими в задаче Атака

  — отдельный лидерборд с 10 защищающимся в задаче Защита

- 😇 Лучшие публичные решения **Companion** оглашаем вместе с лидербордами

  Не тормозите!

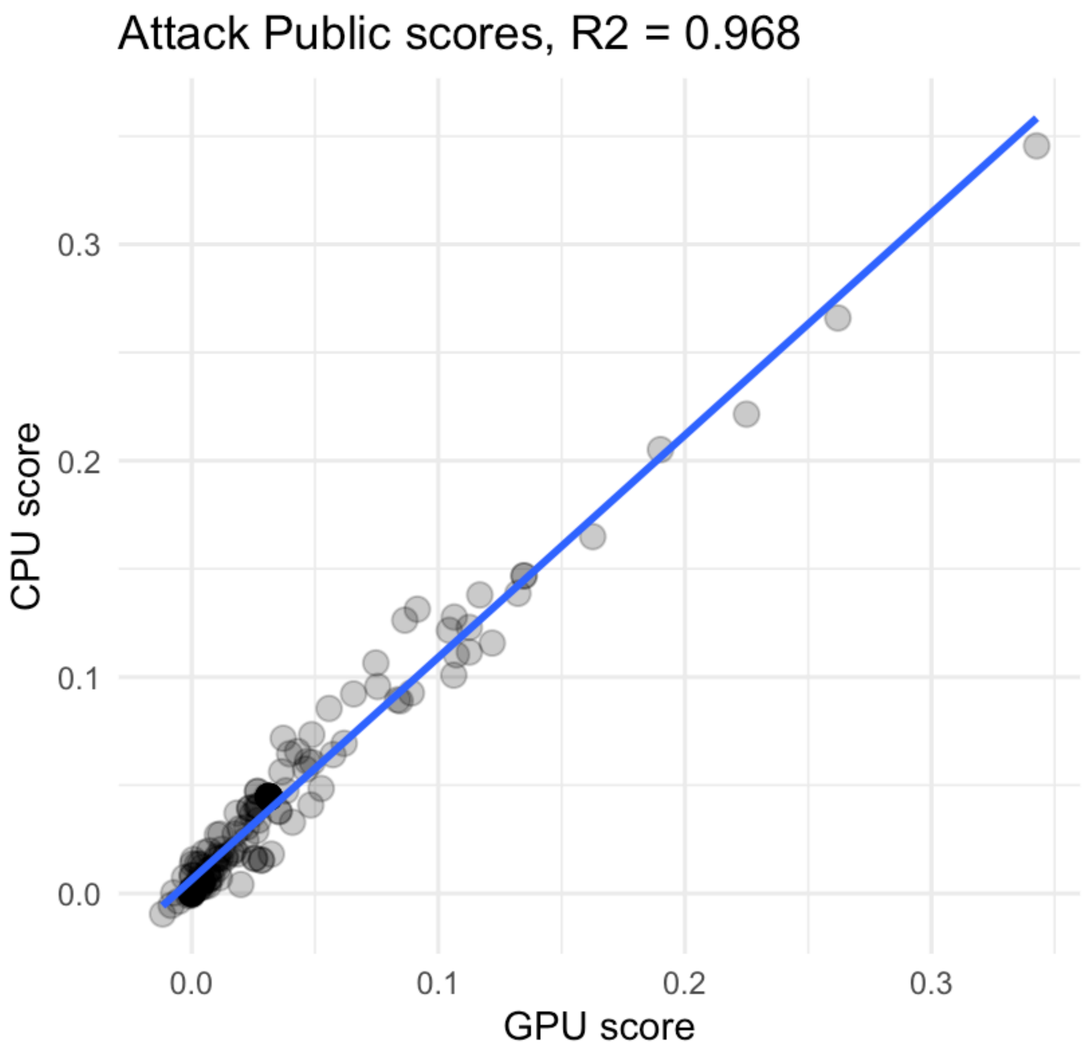
DATA FUSION CONTEST 2023

Часть 4 — bonus

Wildcard про публичные решения

1. Можем на вторую фазу положить
   - Другие атаки для защиты
   - Другую модель для атаки
2. ...При паре условий:
   - Только публичные решения Фазы 1
   - Решения должны быть не хуже организаторских
   - За это проголосуют в telegram чате (да/нет)
   - В чате выберут 1 модель и 1..3 атаки

TIL: GPU vs CPU inference



Спасибо!

Telegram:
Data Fusion Contest 2023 Chat