PWS Cup 2023
コンテストイメージ

PWS 2023 実行委員会

※本資料は，PWS Cup 2023 への参加をご検討中の方に向けて，コンテスト内容をイメージいただくことを目的に，2023/8/17 時点での実施内容の案を公開するものです．実際のコンテストは本資料の内容とは異なるものとなる場合があります．

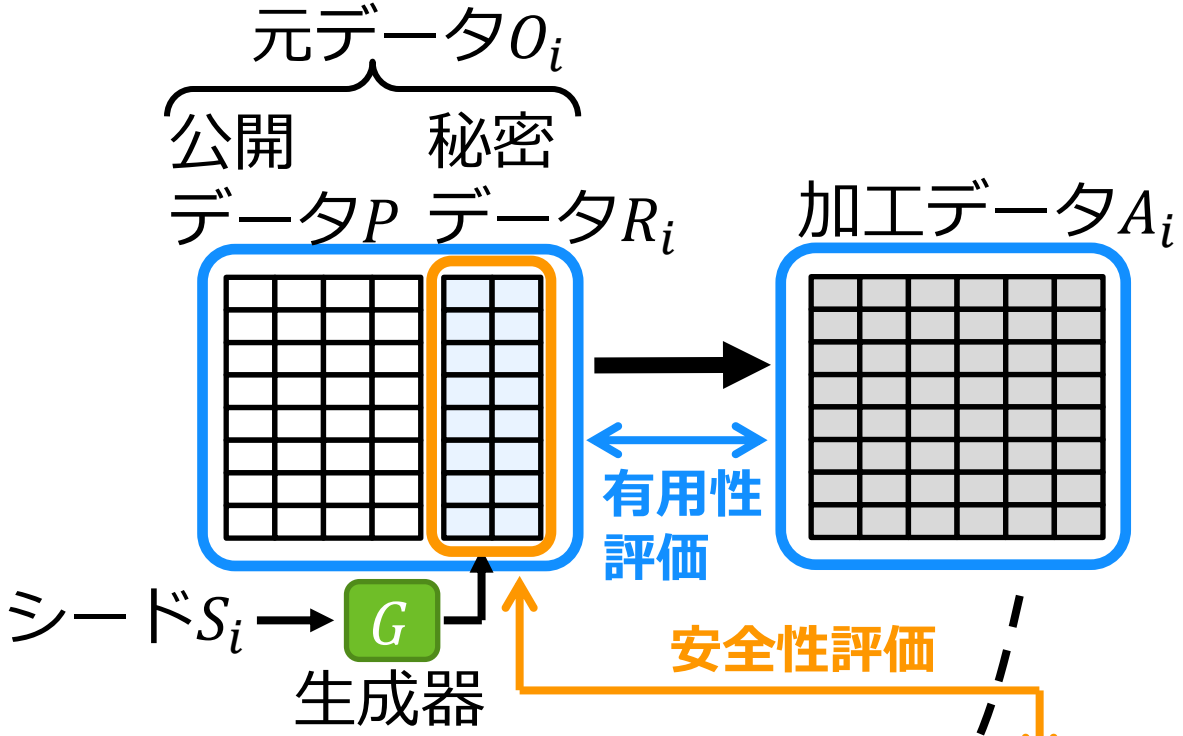
コンテスト実施の目的

- **属性推定**による攻撃を想定したときに
  どこまで**安全性**と**有用性**を保った加工ができるのか
  をよく見てみたい

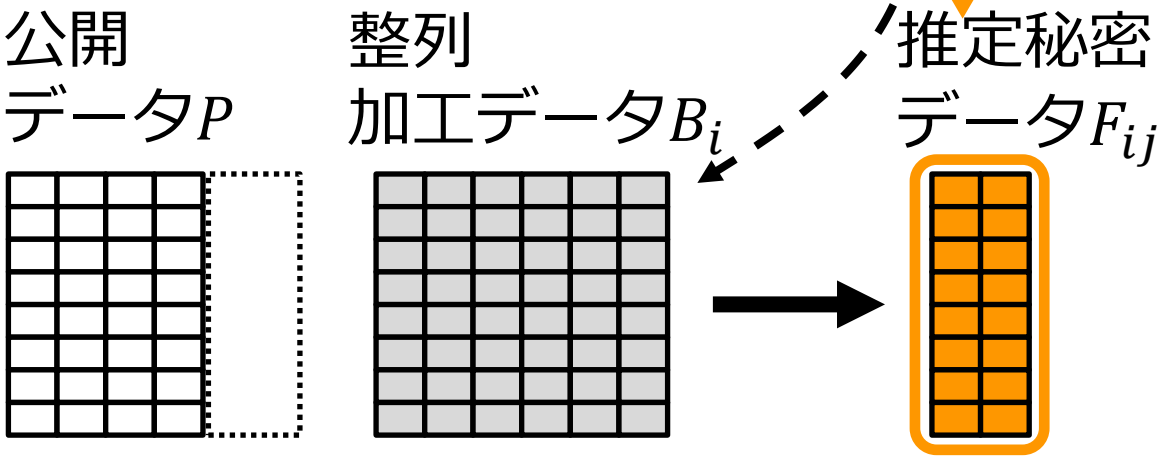
➔ 制限少な目，簡素な設計の，
匿名化・属性推定コンテスト

コンテストの流れ

1. **加工段階**
   (iが加工)



2. **攻撃段階**
   (jがiを攻撃)

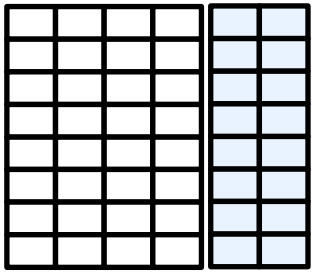


有用性評価

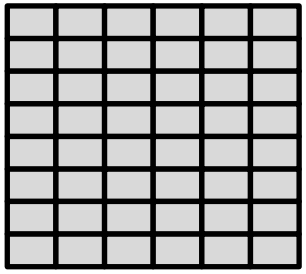
加工データA_i内の加工されていないセルの割合u_iで評価(大きいほどよい)

u_i := (O_iとA_iで値が変わらなかった要素の割合)

元データO_i



加工データA_i

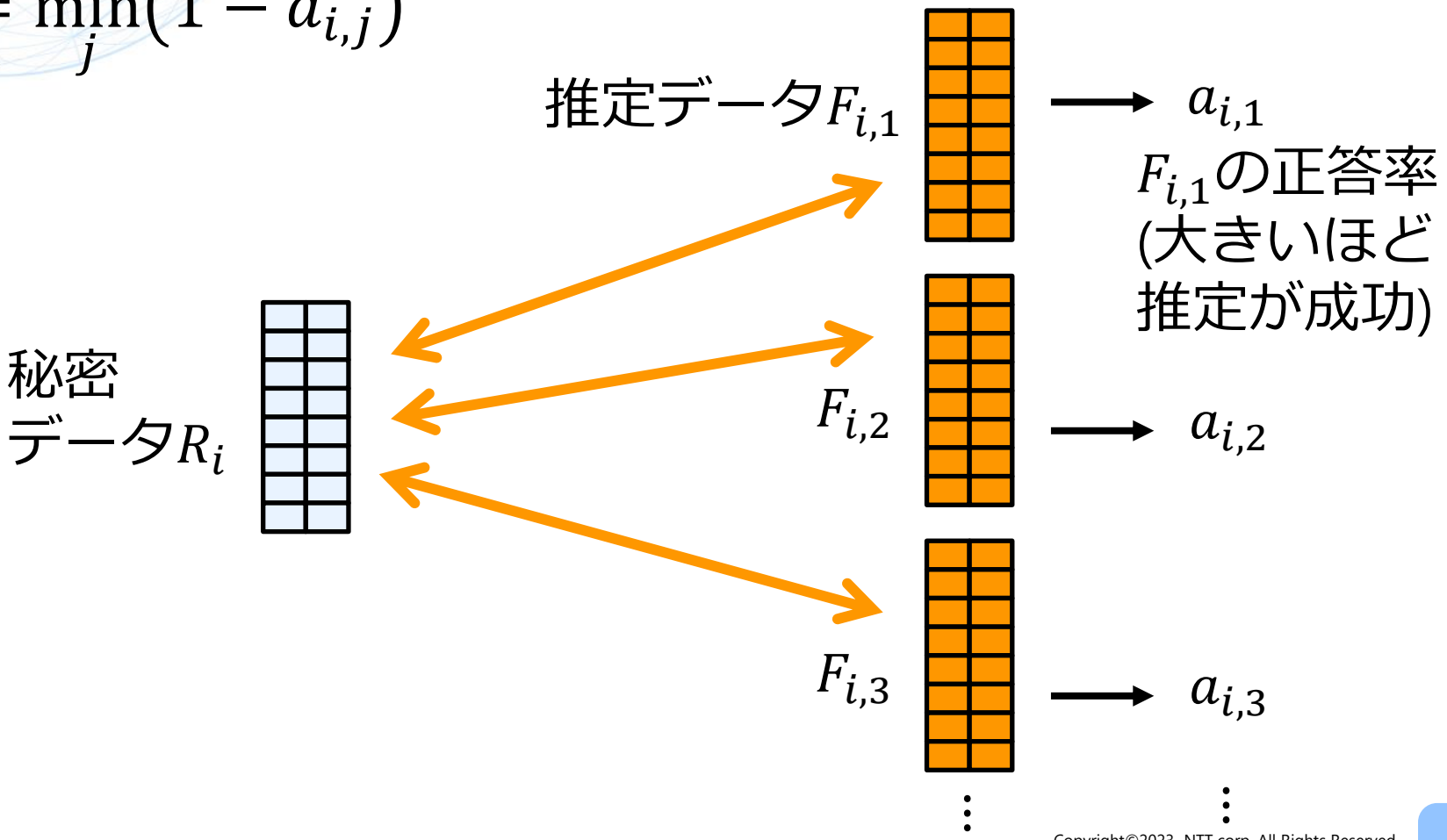


安全性評価

加工データA_iの最小誤答率 s_i で評価(大きいほどよい)

$a_{i,j}$:= (R_iとF_iの一致した要素の割合)

$$s_i := \min_j (1 - a_{i,j})$$



データセット

- 1行1顧客の表
- 左右2つの部分表により構成
  - 公開データP: 出題者が作成し，公開
  - 秘密データR_i: 各チームで選んだ乱数のシードS_iから
    セルごとに独立に，一様ランダムに生成
- 各セルはc値のカテゴリ値$\in \{0,1,\cdots,c-1\}$

公開データP

年代，居住エリア，…

5,8,9,5,0,0
1,7,6,9,2,4
5,2,4,2,4,7
7,9,1,7,0,6
9,9,7,6,9,1

秘密データR_i

(離散化した)体重，年収，…

8,8,6,2,8,7,2,1,5,4
4,5,7,3,6,4,3,7,6,1
3,5,8,4,6,3,9,2,0,4
2,4,1,7,8,2,9,8,7,1
6,8,5,9,9,9,3,0,0,2

パラメータ設定

- 行数 $n = 10^5$
- 公開データ列数 $m_P = 6$
- 秘密データ列数 $m_R = 10$
- カテゴリ数 $c = 10$