

Técnicas de anonimidad

Se busca evitar que un dataset revele las identidades de los usuarios almacenados en este. Los datos que se quieren mantener privados son números como el del DNI, la Seguridad Social, el teléfono o el de una tarjeta de crédito, entre otros.

- **Data Masking:** Consiste en ocultar o alterar valores en un dataset para que el valor original no pueda ser recuperado. Estas modificaciones del valor original se pueden hacer con cifrado, desordenado, diccionarios de sustitución, sustitución de caracteres....
- **Pseudoanonimización:** Reemplaza los identificadores personales por pseudónimos o identificadores artificiales. Normalmente un sistema de enlazado interno se mantiene entre los datos originales y los mostrados, de forma que puedan ser revertidos usando la información apropiada.
- **Generalización:** reemplaza unos valores específicos con valores más amplios o categorías. Normalmente requiere una cantidad de datos enorme para asegurar que los grupos no sean lo suficientemente ambiguos como para ser inútiles.
- **Data Swapping:** Se cambian valores de las filas en las mismas columnas.
- **Data Perturbation:** Sazona los datos con ruido aleatorio y redondeando, mientras que trata de mantener las propiedades para permitir análisis de datos.
- **Synthetic Data:** En vez de publicar datos reales, se toma el dataset original y se pasa por un modelo generativo de forma que este suelte datos inventados pero que encajen con lo que hay en la base de datos.

K-Anonimidad

Un dataset es K-anónimo cuando hay al menos K registros diferentes que comparten los mismos Quasi identificadores. Para cualquier registro (fila), hay al menos k-1 otros registros que comparten el mismo set de atributos que podrían ser usados para identificar cualquiera de ellos de forma única. El valor K se usa para medir la privacidad.

From:

<https://www.knoppia.net/> - **Knoppia**

Permanent link:

<https://www.knoppia.net/doku.php?id=pan:anontec&rev=1729697347>

Last update: **2024/10/23 15:29**

