1)

Netflix proporcionó un amplio conjunto de datos de clasificaciones de usuarios como datos de capacitación, donde los usuarios y las películas fueron reemplazados por identificaciones numéricas

Credit Cards: Las empresas, las instalaciones médicas y las agencias gubernamentales a menudo afirman que desinfectar los datos que almacenan puede mantener la confidencialidad del cliente, paciente o constituyente en caso de pérdida o robo de información. Utilizando un enfoque a veces denominado "desidentificación", limpian los campos de datos que podrían revelar la identidad de uno, confiando en el anonimato como una forma de evitar las obligaciones legales de informar al público cuando hay una fuga o violación de la base de datos.

Health: Para superar las limitaciones que presentan las técnicas de anonimización convencionales, se han propuesto otras técnicas para la anonimización con el objetivo de proporcionar garantías científicas (demostrables) de anonimato (k-Anonymity)

2)

Case1:  Dos investigadores de la Universidad de Texas lograron vincular algunos de los datos "anonimizados" de Netflix con datos externos, revelando así todo el historial de visualización de Netflex para ciertas personas.

Case2: Los investigadores analizaron tres meses de registros de tarjetas de crédito "simplemente anonimizados", sin nombres, números de cuenta u otros identificadores obvios, para 1,1 millones de personas comprando en 10,000 tiendas repartidas por un país no identificado. Esta información fue proporcionada por un banco que opera en ese país. En el conjunto de datos que analizaron se incluyeron los nombres y ubicaciones de las tiendas donde se realizaron las compras, los días en que se realizaron y los montos de las compras. Utilizando tanto la información de la tarjeta de crédito como de la transacción, los investigadores identificaron al 90 por ciento de las personas en el conjunto de datos. Cuando agregaron los precios exactos de las transacciones a la mezcla, aumentaron su capacidad para volver a identificar registros anónimos en un 22 por ciento. Los investigadores descubrieron que podían identificar a las personas incluso si solo conocían su ubicación general o un marco de tiempo durante el cual las personas compraban.

Case3: Los ataques son posibles, en caso de que la diversidad de los atributos no anónimos sea baja. En el ejemplo, todos los hombres con edades comprendidas entre 60 y 69 sufren enfermedades relacionadas con el corazón. Entonces, si un atacante sabe que Bill está incluido en el conjunto de datos, se entera directamente de su enfermedad

3)

PII significa "Personally Identifiable Information" es la información que las organizaciones pueden usar por sí mismas o con otra información para identificar, contactar o ubicar a una sola persona, o para identificar a un individuo en contexto.

4)

Se define la K en  **k-anonymity** como el numero de registros que pueden identificar a una persona en un conjunto de datos.

5)

Cuando se realizan análisis sobre los datos desinfectados, no se produce interacción con los datos originales. En configuraciones interactivas, el analista consulta directamente los datos originales. Si queremos evitar que la PII se filtre al analista en tal entorno, los resultados devueltos deben desinfectarse. Esto se logra típicamente distorsionando los resultados agregando ruido.