Reto Privacidad y Seguridad de los Datos

- Antes de subir los datos a cualquier repositorio, discutan a profundidad con el socio formador sobre la naturaleza de los datos con los que van a trabajar. Es importante entender las implicaciones legales y de seguridad que vienen asociadas a los datos con los que estarán trabajando.

Anonimizar datos:

1. ¿Qué información queremos obtener de dataset?
2. Eliminar los datos que no son relevantes para lo que queremos sacar.
3. Aplicar funciones hash
    - Verifica que los datos que recibes están anonimizados, es decir que no se pueda rastrear información personal o sensible a una persona o producto específico a través del data set. Si los datos ya están anonimizados, describe cuáles fueron los atributos y las razones por las que se tienen que enmascarar.

Primero vamos a comprobar que los datos recibidos están completamente anonimizados para ello observamos cómo se nos entregaron:

PHONE_ID	timestamp	bts_id	lat	lon
668f7c17a62c937a75f762c7198a7fc98ed4e0e0c64ce0...	2021-01-01T17:22:55.000-03:00	CEMG1	-33.3913	-70.6222
780fc36e9a2bc99de12adb740e5e82b3cabba75c1ecd23...	2021-01-01T17:10:19.000-03:00	CEMG1	-33.3913	-70.6222
b3c52936d4f8494dae9d1158ce76951e62413d511f5fe2...	2021-01-01T00:35:04.000-03:00	CEMG1	-33.3913	-70.6222
a4ab622fe4c0de513c389ab475cee4ad5b5d27e07e32d9...	2021-01-01T17:22:38.000-03:00	CEMG1	-33.3913	-70.6222
77d8edaa34e7ac318ef33541957e9f33826dff24217636...	2021-01-01T17:11:12.000-03:00	CEMG1	-33.3913	-70.6222
84124a9088ce1f56e175f5d7d0888e5a3a1a9073c2687c...	2021-01-01T17:10:27.000-03:00	CEMG1	-33.3913	-70.6222

Los datos los que tenemos son:

- PHONE_ID: son los números de teléfono, anonimizados con algoritmo hash de 64 caracteres.
- timestamp: son las “fechas” de la conexión de cada teléfono a alguna antena, previamente transformadas por el socio.
- bts_id: son las etiquetas de las torres de antenas que tiene cada comuna en Santiago de Chile.
- lat: Latitud en la que se encuentra la antena
- lon: Longitud exacta en donde se encuentra la antena.

Además, varios de los datos fueron eliminados del dataset original, como género, lugares diferentes a la zona metropolitana de Santiago de Chile, etc.

En cuanto a la disociación de los datos, por cómo se encuentran, se podría intuir que están disociados, ya que no hay forma de saber el número de teléfono del usuario, ni a qué usuario pertenece cada número, pero, realmente con estos datos, podemos saber cómo se movió un usuario durante el dia, donde pasó la noche, la mañana y la tarde, con lo que si contáramos con más días en el dataset podríamos encontrar el patrón de movimiento de algún usuario e intuir la zona donde vive, trabaja/estudia y pasa su tarde.

- Consulta la normativa actual de la industria a la que esté sujeto el socio formador e investiga en reportes técnicos, artículos o foros cuales son los pasos comunes que se toman para garantizar la privacidad de los datos en dicha industria.

Como nos encontramos en México, pero los datos que se nos proporcionaron son de Chile será necesario seguir ambas normativas además de las leyes Internacionales.

Recordemos que estamos haciendo uso de datos personales en poder de una empresa, por lo que debemos acatar la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP). Esto dado que nosotros, quienes gestionamos la información, nos encontramos en México. Además, si bien los datos son administrados por un procesador de cualquier lugar del mundo, en este caso están siendo almacenados en Google Drive y trabajados en Google Colab (cuya infraestructura está localizada en Estados Unidos), pero las operaciones son realizadas a nombre de un controlador localizado en México, siendo el Tecnológico de Monterrey.

Dentro de la LFPDPPP y en base a los objetivos de nuestro proyecto, debemos seguir los siguientes artículos:

- Artículo 6: Los responsables en el tratamiento de datos personales, deberán observar los principios de licitud, consentimiento, información, calidad, finalidad, lealtad, proporcionalidad y responsabilidad, previstos en la Ley.
- Artículo 10 según el índice III: No será necesario el consentimiento para el tratamiento de los datos personales cuando los datos personales se sometan a un procedimiento previo de disociación.
- Artículo 12: El tratamiento de datos personales deberá limitarse al cumplimiento de las finalidades previstas en el aviso de privacidad. Si el responsable pretende tratar los datos para un fin distinto que no resulte compatible o análogo a los fines establecidos en aviso de privacidad, se requerirá obtener nuevamente el consentimiento del titular.

- Artículo 19: Todo responsable que lleve a cabo tratamiento de datos personales deberá establecer y mantener medidas de seguridad administrativas, técnicas y físicas que permitan proteger los datos personales contra daño, pérdida, alteración, destrucción o el uso, acceso o tratamiento no autorizado.
- Artículo 20: Las vulneraciones de seguridad ocurridas en cualquier fase del tratamiento que afecten de forma significativa los derechos patrimoniales o morales de los titulares, serán informadas de forma inmediata por el responsable al titular, a fin de que este último pueda tomar las medidas correspondientes a la defensa de sus derechos.
- Artículo 21: El responsable o terceros que intervengan en cualquier fase del tratamiento de datos personales deberán guardar confidencialidad respecto de éstos, obligación que subsistirá aun después de finalizar sus relaciones con el titular o, en su caso, con el responsable.

Para Chile debemos seguir las siguientes leyes:

- Ley N° 19628 (PROTECCIÓN DE DATOS DE CARÁCTER PERSONAL) - sobre protección de la vida privada, haciendo un especial énfasis en el artículo 1° sobre el tratamiento y regulación de datos, 2° subsecciones a) Almacenamiento de datos, f) Identificación de datos personales, g) Datos sensibles, i) Fuentes accesibles al público y o) Tratamiento de datos, 4° restricciones para el tratamiento de datos personales, 7° mantener discreción de los datos que se están manejando, 8° sobre el mandato para el tratamiento de datos y 9° tener fines establecidos para el uso de datos. [1]
- Ley N° 20285 (SOBRE ACCESO A LA INFORMACIÓN PÚBLICA) - para poder acceder a la información pública. Teniendo cuidado únicamente con datos que nos puedan proporcionar información personal. [2]

No obstante es importante tomar en cuenta las siguientes principios y leyes internacionales:

- Artículo 12 (Declaración Universal de Derechos Humanos): Nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques, ya sean de manera física o digital [4].

Los datos serán manejados para su procesamiento en México, por lo cual hay que apegarse a las leyes federales y nacionales para su manipulación:

- Ley Federal de Protección de Datos Personales en Posesión de los Particulares de 2010, las entidades deben seguir los siguientes principios descritos en el artículo 6 para garantizar la protección de datos personales:
    - **Licitud**: deben cumplirse todas las regulaciones impuestas en esta ley.
    - **Consentimiento**: al momento de tratar datos personales, las empresas tienen que obtener el consentimiento expreso e informado en relación con los fines de tratamiento definidos.
    - **Información**: siempre debe notificarse al titular cómo se administrarán sus datos, mediante un aviso de privacidad.
    - **Calidad**: los datos tienen que ser exactos, completos, pertinentes, correctos y actualizados en relación con los fines de su gestión.
    - **Finalidad**: la gestión de datos debe poseer un objetivo que será establecido en el aviso de privacidad.
    - **Lealtad**: en todo momento se debe priorizar el interés del titular de los datos, así como garantizar que estos no sean recabados o tratados por medios fraudulentos. La seguridad de los clientes y la reputación de la marca dependen de ello.
    - **Proporcionalidad**: Los datos recabados y tratados deben ser solo aquellos que se consideran necesarios, adecuados y pertinentes para el cumplimiento de las finalidades. Se debe evitar pedir datos que a larga puedan traducirse en discriminación o racismo.
    - **Responsabilidad**: las empresas son 100% responsables de los datos personales que recolectan, y deben hacerse cargo del tratamiento que hagan de esa información.

- Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados de 2017:
    - Los sindicatos y cualquier otra persona física o moral que reciba y ejerza recursos públicos o realice actos de autoridad en el ámbito federal, estatal y municipal serán responsables de los datos personales, de conformidad con la normatividad aplicable para la protección de datos personales en posesión de los particulares.

- Establece un proceso claro sobre cómo se puede trabajar con el set de datos y especifica aspectos como: dónde se puede almacenar, en que tipo de redes puede estar, quien los puede ver y cuales son los documentos o normas que se deben de firmar antes de poder acceder a los datos.

Para mantener la seguridad de los datos, estos se almacenarán únicamente en un Google drive privado, donde solo se dará acceso a los miembros del equipo, profesores del curso y algunos miembros de otros equipos. Drive cuenta con su propio sistema de seguridad, el cual detecta y protege de amenazas.[3]
En cuanto a los códigos del proceso se guardaran en Github, en un repositorio público, por lo que no subiremos los datos brindados por cuestiones de seguridad.
Igualmente los datos se almacenarán de forma local en las computadoras de los miembros del equipo.

- Implementen un mecanismo o utiliza una herramienta que les permita establecer registros sobre quien y cuando tuvo acceso a los datos y bajo qué esquema. Estos registros los deberán integrar a su reporte como parte de la evidencia de final de módulo.

En nuestro caso, como tal, no contamos con un log para saber quienes han accedido a los datos brindados, pero, al tener los datos resguardados únicamente en drive y local, con esto aseguramos que terceros no tengan acceso, los únicos con acceso a nuestro drive son miembros del equipo, profesores y algunos compañeros de la concentración.

Bibliografía:

1. Biblioteca del Congreso Nacional de Chile. LEY 19628 SOBRE PROTECCIÓN DE LA VIDA PRIVADA. Retrieved November 2, 2022, from https://www.bcn.cl/leychile/navegar?idNorma=141599
2. Biblioteca del Congreso Nacional de Chile. LEY 20285 SOBRE ACCESO A LA INFORMACIÓN PÚBLICA. Retrieved November 2, 2022, from https://www.bcn.cl/leychile/navegar?idNorma=276363
3. *De qué manera drive protege Tu Privacidad y te da el control - ayuda De Drive* (no date) *Google*. Google. Available at: https://support.google.com/drive/answer/10375054?hl=es-419&dark=1 (Accessed: November 2, 2022).
4. Declaración Universal versión comentada. CORTEIDH. Retrieved November 2, 2022, from https://www.corteidh.or.cr/tablas/28141.pdf
5. Ley Federal de Protección de Datos. Docusign. Retrieved November 2, 2022, from

https://www.docusign.mx/blog/proteccion-de-datos-personales#:~:text=La%20Ley%20Federal%20de%20Protecci%C3%B3n%20de%20Datos%20Personales%20en%20Posesi%C3%B3n,en%20poder%20de%20las%20empresas.

6. Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados. Camara de diputados. Retrieved November 2, 2022, from https://www.diputados.gob.mx/LeyesBiblio/pdf/LGPDPPSO.pdf