TECNOLÓGICO DE MONTERREY®

Inteligencia artificial avanzada para la ciencia de datos II

Reto

Reto: Privacidad y Seguridad de datos

Diego Arturo Padilla Domínguez - A01552594

Keyuan Zhao - A01366831

Carolina Herrera Martínez - A01411547

Cutberto Arizabalo Nava - A01411431

Jose Pablo Cobos Austria - A01274631

Campus Querétaro

1 de noviembre de 2022

Privacidad y Seguridad de datos

Verificación de anonimización

En este punto hablaremos sobre dos dataset importantes, los que se nos otorgaron al inicio del reto por parte del socio formador y los que tenemos actualmente. Empezando por los datos que nos dieron la inició, estos datos si cumplen con la anonimización requerida, ya que el socio formador enmascarado los datos mas que serian los mas sensibles a ser utilizados para identificar a los individuos registrados en el dataset, como puede ser datos como el nombre y apellido o fechas de los registros que se puedan relacionar con otros datasets y comprometer a los usuarios de quien se obtuvo la información. Esto lo realizó mediante diferentes métodos para realizar esta anonimización, como lo fue la eliminación de columnas que no son relevantes para la solución del reto, formatear la fecha de los registros y finalmente mediante funciones hash anonimizar los nombres o identificadores de los usuarios registrados.

Ahora sobre el dataset que nosotros utilizamos para nuestra matriz de viajes y aplicar el modelo de machine learning, nos hemos asegurado que los datos estén lo suficientemente anonimizados ya que este dataset solamente contiene las comunas, las distancias, y la cantidad de viajes que se hizo entre cada una con las horas correspondientes, que especificamente este ultimo podria anonimizar aun màs si se aplicara el método de rango y las horas se “redondeará” a la hora más cercana

Investigación de documentación técnica

En cuanto al procedimiento que se debe seguir para garantizar la privacidad de los datos, podemos encontrar documentos como la ***Ley Federal de Protección de Datos Personales en Posesión de los Particulares*** y guías emitidas por el ***Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales*** *("INAI").*

1. ¿Mi organización trata datos personales en el ejercicio de sus actividades cotidianas?

2. ¿Qué figura tiene mi organización? ¿es responsable o encargado?

3. ¿Qué tipo de datos personales trata mi organización?

4. ¿Alguno de estos datos personales son patrimoniales, financieros o sensibles?

5. ¿De dónde se obtienen los datos personales?

6. ¿Qué persona, área o departamento de la organización trata los datos personales?

7. ¿Para qué fines se tratan los datos personales?

8. ¿Se comunican datos personales a encargados?

9. ¿Se comunican datos personales a personas físicas o morales que no sean encargados? ¿a

quién y para qué se comunican los datos?

10. ¿Dónde se almacenan los datos personales?

11. ¿Por cuánto tiempo se conservan los datos personales?

12. ¿Cómo se borran o eliminan los datos personales?

Procedimientos y reglas para almacenamiento y acceso de los datos

El dataset compartido por parte del socio formador está almacenado de forma local y virtual (en la nube), y sólo las personas autorizadas pueden acceder.

- **Local**

Todos los integrantes del equipo tienen almacenado el dataset de forma local para realizar cualquier tipo de pruebas necesarias para la implementación del modelo, ya que de esta forma se agiliza el tiempo.

- **Virtual**

Drive: lo cual fue originalmente compartido por parte del socio formador hacia nosotros y sólo las personas que tiene el enlace pueden acceder. Además, creamos una unidad compartida para guardar el dataset y otros códigos de procesamiento, esto sólo los integrantes del equipo pueden tener acceso hacia ello.

Servidor en la nube: sólo las personas que tienen la llave privada, el usuario y la contraseña pueden acceder.

El mecanismo para ver el registro de los accesos a nuestros códigos y al dataset, se utiliza por default que trae las plataformas. En Drive se puede visualizar quienes tienen acceso a las

carpetas compartidas y qué reglas tienen cada una de ellas (leer, comentar o editar). En el servidor no hay forma de visualizar quienes accedieron a los códigos y al dataset, por lo tanto establecimos sólo a las personas tanga la llave privada, el usuario y la contraseña para acceder.