



Tecnológico de Monterrey
Campus Querétaro

Inteligencia artificial avanzada para la ciencia de datos II
Grupo 501

Momento de Retroalimentación

Reto Privacidad y Seguridad de los Datos

Maestro

Benjamín Valdés Aguirre

Presenta

Adrián Galván Díaz A01704076

28.10.2024

Introducción

La seguridad de los datos es un pilar fundamental dentro de proyectos de machine learning, especialmente cuando trabajamos con información que podría considerarse sensible, como imágenes de personas o datos personales de las mismas. Este trabajo busca no solo desarrollar modelos que optimicen el uso de las camas de las vacas para mejorar su producción de leche, sino también garantizar que toda la información manejada esté protegida bajo los principios éticos y las normativas legales de privacidad. A través de este reporte, se documentarán los pasos esenciales para asegurar que los datos utilizados estén anonimizados, además de detallar el proceso de gestión y acceso seguro a estos datos. Así, nuestro objetivo es generar información valiosa para el socio formador sin comprometer la privacidad ni la seguridad de la información con la que trabajamos.

Anonimización de Datos | Normativas y Técnicas

Para asegurar la privacidad de los datos en nuestro proyecto, el primer paso es verificar si los datos que manejamos están correctamente anonimizados. Este proceso es crucial, ya que el trabajo con imágenes de camas de vacas podría incluir información sensible que podría ayudar a saber la ubicación del rancho o identificar la identidad de un trabajador.

Después de revisar los datos en equipo, nos dimos cuenta que las imágenes no cuentan con información sensible como datos geolocalizables o metadatos visibles. A pesar de eso, cualquier detalle particular del entorno (marcas, colores específicos, equipos únicos) podría permitir una identificación indirecta del rancho. Sin embargo este tampoco es el caso, las imágenes que se nos proporcionaron son solo de las camas de las vacas sin nada que pueda ser identificable.

Una de las últimas verificaciones que teníamos que hacer era ver si aparecían personas en las imágenes. En los datos que revisamos como equipo obtuvimos 1 foto donde aparecía una persona de espaldas sin que se le viera la cara en 5000 fotos. En cuanto a la presencia de personas en las fotos, aunque estén de espaldas y no se vean sus rostros, la inclusión de personas en cualquier dataset que se utiliza para machine learning puede ser un aspecto sensible desde la perspectiva de la privacidad.

En México, la protección de datos personales está regulada principalmente por la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) y la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO). Ambas leyes establecen que cualquier información que permita identificar a una persona debe tratarse de manera confidencial y bajo condiciones de seguridad, especialmente si esta información se recopila para análisis o modelos, como es nuestro caso.

La LFPDPPP considera datos personales cualquier información que identifique directa o indirectamente a una persona. En nuestro caso, si las imágenes capturan personas, incluso si están de espaldas, pueden incluir rasgos únicos que permitan una identificación, lo cual entra en el ámbito de protección de esta ley. Algunas técnicas que podemos usar para evitar caer en esta ley son las siguientes:

- Aleatorización:

Adición de Ruido: Esta técnica puede ser útil para distorsionar detalles de la imagen (como características del entorno o de personas que aparecen de espaldas) sin afectar la información esencial de las vacas o del suelo.

- Seudonimización:

Cifrado por Hash: Si las imágenes contienen algún tipo de etiqueta o metadata que podría vincularse a una identidad (como identificadores de dispositivos, identificadores de uniformes, etc), el cifrado por hash puede proteger estos datos.

Una de las leyes que también puede entrar en juego es el artículo 16 de la Constitución Política de los Estados Unidos Mexicanos. Esta establece el derecho a la protección de los datos personales, otorgando a los individuos la posibilidad de acceder, rectificar, cancelar, y oponerse al uso de sus datos personales. Este derecho es clave para cualquier tratamiento de información que incluya datos identificables de personas, como en el caso de fotografías donde podría aparecer personal de un rancho. Es por esa razón que aplicar un desenfoque o una anonimización como las que se listaron arriba son esenciales para el uso de datos.

Registro de Acceso a los Datos | Proceso de Trabajo con el Conjunto de Datos

Para asegurar la trazabilidad y seguridad en el manejo de los datos, es necesario también establecer un sistema de registro que documente cada acceso al conjunto de datos. Este registro no solo cumple con normativas de seguridad, sino que también permite un monitoreo continuo de las acciones realizadas con los datos, minimizando el riesgo de accesos no autorizados o de uso indebido. Para gestionar estos registros, se puede implementar una herramienta o sistema que capture automáticamente la información de acceso, incluyendo el usuario, la fecha y la hora en la que se realizó. La ISO/IEC 27001 enfatiza la necesidad de controlar el acceso a la información mediante mecanismos de autenticación y autorización (Cláusula A.9.1). Implementar una herramienta de registro permite gestionar y monitorear quién accede a los datos, cuándo lo hace y con qué propósito.

Por otro lado, la norma establece que toda actividad en el sistema debe ser registrable y auditable (Cláusula A.12.4) por lo que mantener evidencia de acceso asegura la trazabilidad y todos los datos necesarios pueden ser registrados por dicha herramienta.

Proceso de Verificación de Datos Anonimizados (definido por el equipo)

Revisión Previa a la Carga: Las imágenes serán revisadas antes de su carga para identificar cualquier elemento que pueda comprometer la privacidad de las personas.

Difuminado Selectivo: En caso de detectar rostros, se aplicará un difuminado a dichas áreas antes de almacenarlas en la base de datos del proyecto.

Control de Acceso: Los datos serán accesibles únicamente para el equipo autorizado, cumpliendo con los estándares de seguridad establecidos en la normativa mexicana e internacional (Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP) e ISO/IEC 27001).

Procedimientos de Control de Acceso y Seguimiento de Proceso (definido por el equipo)

Para garantizar que el acceso a los datos esté restringido y en cumplimiento con la normativa, se implementarán los siguientes controles:

Roles y Permisos: El acceso a los datos de imágenes estará limitado a los miembros del equipo de desarrollo y análisis, quienes deberán seguir procesos específicos de autenticación.

Registro de Acceso: Se mantendrán registros detallados de quién accede a los datos y en qué momento, permitiendo un seguimiento claro y transparente de la manipulación de los datos.

Autorizaciones y Validación de Manejo de Datos: Antes de utilizar las imágenes, se requerirá una aprobación de cumplimiento de normas, validando que las prácticas de anonimización y seguridad estén correctamente aplicadas.

Referencias

ANEXO_1- Marco legal protecci_n de datos personales. (s. f.).

https://www.gob.mx/cms/uploads/attachment/file/525742/ANEXO_1- Marco legal protecci_n de_datos_personales.pdf

De Agricultura y Desarrollo Rural, S. (s. f.). Protección de datos personales. gob.mx.

<https://www.gob.mx/agricultura/acciones-y-programas/proteccion-de-datos-personales-282241>

De la Función Pública, S. (s. f.). Normatividad en materia de Protección de Datos Personales. gob.mx.

<https://www.gob.mx/sfp/documentos/normatividad-en-materia-de-acceso-a-la-informacion-y-proteccion-de-datos-personales-nuevo?state=published>

Normativa y legislación en PDP – Marco Internacional de Competencias de Protección de Datos Personales para Estudiantes. (s. f.).

https://micrositios.inai.org.mx/marcocompetencias/?page_id=370

Marco normativo. (s. f.). Sistema Nacional de Transparencia Acceso a la Información Pública y Protección de Datos Personales

<https://proyectos.inai.org.mx/pronadatos/index.php/inicio/informacion/marconormativo>

Solutions, G. (2023, 22 septiembre). ¿Qué es la norma ISO 27001 y para qué sirve? GlobalSuite Solutions.

<https://www.globalsuitesolutions.com/es/que-es-la-norma-iso-27001-y-para-que-sirve/#:~:text=La%20norma%20ISO%2027001%20es,y%20disponibilidad%20de%20la%20informaci%C3%B3n.>

Eee. (2019, 17 septiembre). Cómo gestionar los controles de acceso según ISO 27001. Escuela Europea de Excelencia.

<https://www.escuelaeuropeaexcelencia.com/2019/09/como-gestionar-los-controles-de-acceso-segun-iso-27001/>

López, A. (s. f.). Anexo 12.

https://www.iso27000.es/iso27002_12.html#:~:text=Controles%20del%20riesgo,-12.4.1%20Registro&text=12.4.4%20Sincronizaci%C3%B3n%20de%20relojes,de%20sincronizaci%C3%B3n%20%C3%BAnica%20de%20referencia.