Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Querétaro

Momento de Retroalimentación

Reto Privacidad y Seguridad de los Datos

Autor:

A01368818 Joel Sánchez Olvera

TC3007C

Inteligencia artificial avanzada para la ciencia de datos II

Fecha de Entrega:

02 - Noviembre - 2024

Documento de Privacidad y Seguridad en el Análisis de Imágenes de Ganado para Detectar objetos y el estado del ganado

Introducción

Este documento tiene como finalidad establecer las prácticas y normativas necesarias para garantizar la privacidad y seguridad de los datos utilizados en el proyecto de detección de objetos y clasificación del estado de ganado (vacas) a partir de imágenes mediante el uso de aprendizaje profundo.

Contexto del Proyecto

El proyecto se enfoca en la clasificación y detección de objetos en imágenes para evaluar el estado del ganado (vacas) en un entorno definido. La mayoría de las imágenes capturan únicamente a las vacas y su entorno inmediato, es decir las camas donde se encuentran, sin la presencia de personas ni elementos identificables. Sin embargo, en casos excepcionales, podrían aparecer personas en el fondo o en segundo plano de algunas imágenes. Este documento detalla las medidas de anonimización y control de acceso, así como los aspectos legales aplicables al proyecto mencionado.

Consideraciones de Seguridad y Privacidad

Anonimización de Datos

Para garantizar la privacidad en las imágenes, se recomiendan las siguientes técnicas en caso de que en futuras iteraciones se capture información que incluya personas u otros elementos identificables:

Técnica	Descripción	Aplicabilidad
Hashing de Imágenes	Aplicación de hashing para encriptar imágenes durante el almacenamiento.	Recomendado para mejorar la seguridad en almacenamiento.
Tokenización	Asignación de tokens a metadatos sensibles,	No aplicable directamente a imágenes,

	reduciendo riesgos de exposición.	pero útil en descripciones o etiquetas.
Difuminado	Desenfoque de partes específicas de la imagen (rostros u otros elementos) para anonimizar.	Aplicable si se identifican personas en el fondo de las imágenes.
Adición de Ruido	Se agregan variaciones leves para impedir la identificación exacta de objetos o personas.	Útil para suavizar detalles en el fondo.

Estas técnicas se pueden implementar si en futuras iteraciones se amplía la captura de imágenes y se corre el riesgo de incluir elementos que comprometan la privacidad de las personas.

Control de Acceso

El acceso a los datos debe estar estrictamente limitado a personal autorizado y asegurarse mediante la implementación de controles, como autenticación de múltiples factores (MFA), restricciones de permisos según el rol y auditorías periódicas. Estos mecanismos contribuyen a minimizar el riesgo de acceso no autorizado, asegurando que solo aquellos involucrados en el proyecto puedan manipular y analizar las imágenes.

Almacenamiento Seguro y Manejo de Datos

Para garantizar la seguridad de los datos almacenados, podemos hacer uso de sistemas de almacenamiento que cumplan con estándares de cifrado, como el **AES-256**. Además, es necesario establecer políticas de retención de datos, eliminando cualquier información que no sea esencial para el proyecto una vez que haya sido procesada.

Normativas Mexicanas e Internacionales de Privacidad y Seguridad de los Datos

Normativa en México

Las normativas en México sobre la protección de datos personales están controladas por la **Ley Federal de Protección de Datos Personales en Posesión de los Particulares (LFPDPPP)**.

Esta ley establece la obligación de proteger los datos personales de individuos identificables, y aunque en el proyecto las imágenes contienen principalmente ganado y su entorno, es esencial considerar el cumplimiento de esta normativa en caso de que en futuras iteraciones se capturen elementos o personas identificables que puedan contener o dar a asumir datos personales.

La ley impone sanciones por incumplimiento y exige que se apliquen principios de seguridad, confidencialidad y control de acceso.

Normativa Internacional

Existen normas internacionales relevantes para el manejo seguro de datos en proyectos como este:

- **ISO/IEC 27001:2022**: Establece un marco para la gestión de la seguridad de la información. Cumplir con este estándar ayuda a asegurar que los sistemas de información y los datos almacenados cumplan con prácticas de seguridad de nivel internacional, incluyendo controles para proteger la integridad y disponibilidad de los datos.
- **GDPR (Reglamento General de Protección de Datos de la UE)**: Aunque se aplica principalmente a los datos de ciudadanos europeos, esta regulación es una referencia importante debido a sus estrictas políticas sobre la protección de datos y privacidad. Establece obligaciones claras en relación con el consentimiento informado, el derecho a la eliminación y la seguridad en el almacenamiento y tratamiento de datos personales.

Estas normativas ofrecen una base para garantizar que las prácticas de manejo de datos en el proyecto se alineen con las mejores prácticas de privacidad y seguridad globales.

Responsabilidades y Buenas Prácticas

Evaluación y Revisión Continua

Es esencial realizar evaluaciones periódicas para garantizar que las prácticas de seguridad y privacidad cumplan con las normativas. Estas revisiones también pueden identificar áreas de mejora para fortalecer la protección de los datos y la privacidad en futuras iteraciones del proyecto.

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

1. **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.
2. **Difuminado Selectivo:** En caso de detectar rostros, se aplicará un difuminado a dichas áreas antes de almacenarlas en la base de datos del proyecto.
3. **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 (ver sección 4).

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:

1. **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.
2. **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.
3. **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.

Conclusiones

Podemos concluir que establecemos las pautas de privacidad y seguridad necesarias para manejar los datos del proyecto de análisis del estado del ganado, así como lo que fue definido por el equipo para la anonimización de los datos y el control de acceso y seguimiento de los procesos.

La implementación de técnicas de anonimización y la adherencia a normas de seguridad, como el ISO/IEC 27001:2022 y la LFPDPPP, aseguran que el manejo de datos cumple con estándares de seguridad y privacidad, protegiendo la confidencialidad de cualquier información sensible.

Bitácora de Seguridad de los Datos

Bitácora de Seguridad de Datos						
Actividad	**Descripción**	**Fecha de inicio**	**Fecha de último cambio**	**Responsable de último cambio**	**Personas con acceso**	**Links de acceso**
Google Drive	Carpeta de Google Drive donde el equipo almacena la información necesaria al proyecto	09/10/2024	25/11/2024	Adrián Galván Díaz	Equipo No Name, Profesores	Reto IA
Github	Repositorio de Github donde el equipo almacena el código fuente, documentación necesaria y resultados del proyecto	05/10/2024	24/11/2024	Arturo Cristian Diaz	Equipo No Name, Profesores, Socio Formador	Repositorio
Obtención del dataset original	El socio formador y el profesorado nos dieron acceso a la carpeta de One Drive donde podemos encontrar los datasets para las camas de arena y la fila de ordeño	17/09/2024	17/09/2024	Ivo Ayala	Equipo No Name, Personal de CAETEC, Profesores, Socio Formador	Pictures
Dataset para Modelo Bounding Box	Sets de imágenes que se utilizaron para entrenar, validar y evaluar los modelos de detección de objetos utilizando Tensor Flow 2 y Pytorch	7/10/2024	23/10/2024	Arturo Cristian Diaz	Equipo No Name, Profesores, Socio Formador	Bounding …
Dataset para Modelo Clasificador	Set de imágenes que se utilizó para entrenar, validar y evaluar los modelos de detección de objetos utilizando Tensor Flow 2 y Pytorch	7/10/2024	23/10/2024	Arturo Cristian Diaz	Equipo No Name, Profesores, Socio Formador	Classifier
Dataset para Análisis de Patrones de Arena	Se utilizó el dataset de camas vacías para identificar patrones en la arena.	22/10/2024	24/10/2024	Juan Pablo Cabrera	Equipo No Name, Profesores, Socio Formador	Sand Clas…
Documentación de modelos	Documentos donde se presenta la descripción de los modelos, su justificación, sus parámetros seleccionados y sus resultados	10/10/2024	20/11/2024	Carlos Eduardo Velasco	Equipo No Name, Personal de CAETEC, Profesores, Socio Formador	Documentación Modelos
Resultados de los modelos	Archivos donde se puede acceder a los resultados de cada modelo entrenado para la solución final	24/10/2024	13/11/2024	Juan Pablo Cabrera	Equipo No Name, Personal de CAETEC, Profesores, Socio Formador	Results

Codigo fuente de los modelos	Ultima versión de los códigos fuentes para entrenar, validar y evaluar los modelos para la solución final	6/10/2024	19/11/2024	Arturo Cristian Diaz	Equipo No Name, Personal de CAETEC, Profesores, Socio Formador	**Bounding Box:** Bounding Box **Classifier:** Classifier **Integracion de modelos:** Main
Acceso a la base de datos en la Rasperry Pi	Definición de las personas con acceso a la base de datos generada en la Raspberry Pi	15/11/2024	20/11/2024	Arturo Cristian Diaz	Equipo No Name, Socio Formador, Personal de CAETEC	NA
Acceso al script de la solución desde la Raspberry Pi	Definición de las personas con acceso al script final de la solución generado en la Raspberry Pi	15/11/2024	20/11/2024	Arturo Cristian Diaz	Equipo No Name, Socio Formador	NA

Bitácora de Logs

Logs de Seguridad de Datos			
Google Drive / Github	**Descripción**	**Fecha y Hora**	**Persona que tuvo acceso**
Google Drive	Creación de carpeta	10:11 a.m. 9 oct	Carlos Velasco
Google Drive	Business Understanding	11:44 a.m. 9 oct	Carlos Velasco
Google Drive	Reporte de Descripción de los Datos	6:14 p.m. 9 oct	Joel Sanchez
Google Drive	Reporte de Exploración de los Datos	6:15 p.m. 9 oct	Joel Sanchez
Google Drive	Reportes de Data Understanding	9:27 a.m. 14 oct	Adrian Galvan
Google Drive	Reportes de Data Understanding	9:59 a.m. 16 oct	Juan Pablo Cabrera
Google Drive	Creación carpeta Modelo Bounding Box	9:08 p.m. 20 oct	Arturo Diaz
Google Drive	Tutorial Tensorflow Bounding Box	2:14 p.m. 21 oct	Carlos Velasco
Google Drive	Data Preparation	9:49 a.m. 30 oct	Carlos Velasco
Google Drive	Reporte inicial Análisis de Reporte de Arena	10:44 a.m. 30 oct	Juan Pablo Cabrera
Google Drive	Modificación de Business Understanding	11:37 p.m. 30 oct	Arturo Diaz
Google Drive	Reporte Inicial Classifier	9:45 a.m. 5 nov	Joel Sanchez

Google Drive	Reporte Inicial Bounding Box	6:34 p.m. 10 nov	Carlos Velasco
Google Drive	Editar Reporte Classifier	12:43 p.m. 13 nov	Joel Sanchez
Google Drive	Reporte de Bounding Box TF	9:24 p.m. 19 nov	Carlos Velasco
Google Drive	Creacion Carpeta Evaluación	9:45 a.m. 20 nov	Adrian Galvan

Google Drive	Modificación de Business Understanding	9:49 a.m. 20 nov	Adrian Galvan
Google Drive	Segunda Versión de Clasificador	11:09 a.m. 20 nov	Joel Sanchez
Google Drive	Reestructuración de documentos	12:39 p.m. 20 nov	Carlos Velasco
Google Drive	Guía de iteraciones	2:22 p.m. 20 nov	Joel Sanchez
Google Drive	Reestructuración de documentos de modeling	9:23 p.m. 20 nov	Carlos Velasco
Google Drive	Subir resultados de Sand Classifier	9:54 p.m. 20 nov	Juan Pablo Cabrera
Google Drive	Segunda Versión de Sand Classifier	6:37 p.m. 21 nov	Juan Pablo Cabrera
Google Drive	Modificación de documentos de Modeling	11:09 a.m. 22 nov	Carlos Velasco
Google Drive	Diagrama de Flujo de solución final	11:54 a.m. 22 nov	Carlos Velasco
Google Drive	Modificación guía de iteraciones	2:54 p.m. 22 nov	Juan Pablo Cabrera
Google Drive	Subir resultados de DB	7:01 p.m. 22 nov	Arturo Diaz
Google Drive	Etapa de Entrega	2:44 p.m. 24 nov	Carlos Velasco
Github	Entendimiento de Negocio y creacion de repo	5 oct	Joel Sanchez
Github	Agregar README	9 oct	Arturo Diaz
Github	Primer Modelo Bounding Box	20 oct	Arturo Diaz
Github	Actualizar Bounding Box	24 oct	Arturo Diaz
Github	Calcular coordenadas, centroide y cortar imagenes	1 nov	Carlos Velasco
Github	Clasificador de posiciones	8 nov	Joel Sanchez
Github	Agregar pesos de modelos	13 nov	Arturo Diaz

Github	Implementar Base de Datos	13 nov	Arturo Diaz
Github	Subir clasificadores de arena	15 nov	Juan Pablo Cabrera
Github	Subir documentos de CRISPDM	20 nov	Carlos Velasco
Github	Subir Etapa de Evaluación	22 nov	Adrian Galvan
Github	Actualizar documentos de modeling	25 nov	Carlos Velasco

Logs completos:
Github: Logs