Instituto Tecnológico y de Estudios Superiores de Monterrey

Inteligencia Artificial Avanzada para la Ciencia de Datos

Los concentrados v2

Política de Datos y Acceso

Daniel Queijeiro Albo - A01710441

Diego Alfaro Pinto - A01709971

Diego Isaac Fuentes Juvera - A01705506

Jesus Ramirez Delgado - A01274723

Mauricio Anguiano Juarez - A01703337

Luis Adrián Uribe Cruz - A01783129

Índice

1.0 Objetivo

Esta política establece los lineamientos para la gestión, protección y uso de los datos generados y utilizados en el proyecto de inteligencia artificial para la ganadería de precisión en el CAETEC

2.0 Alcance

Esta política aplica a:

- Datos generados por el sistema DeLaval (robot de ordeño BMC3)
- Imágenes de vacas capturadas por cámaras en corredor de salida
- Datos derivados y modelos generados
- Todo el personal y estudiantes del proyecto

3.0 Clasificación de Datos

Datos Públicos

- Estadísticas agregadas del hato (sin identificación individual)
- Papers académicos y publicaciones
- Modelos open source desarrollados
- Código y metodologías (repositorio GitHub)

Restricciones: Ninguna, pueden compartirse libremente tras finalizar el proyecto

Datos Confidenciales

- Registros individuales por vaca (ID, producción, salud)
- Imágenes originales de las vacas
- Archivos CSV del robot DeLaval
- Métricas de calidad de leche
- Datos que puedan afectar relaciones comerciales

Restricciones: Acceso limitado a equipo del proyecto, profesores y personal autorizado de CAETEC

4.0 Principios Fundamentales

1. **Confidencialidad comercial**
   Los datos son propiedad de CAETEC. No se compartirán fuera del equipo autorizado sin consentimiento expreso.
2. **Transparencia y apertura académica**
   Se promoverá el desarrollo de modelos open source. Los resultados serán compartidos con la comunidad académica respetando la confidencialidad de datos sensibles.
3. **Calidad y confiabilidad**
   Los datos serán validados, limpiados y documentados antes de su uso. Se mantendrán registros de todas las transformaciones realizadas.
4. **Ética y bienestar animal**
   El uso de datos no comprometerá el bienestar del ganado bovino. Todas las decisiones basadas en datos priorizarán la salud y el bienestar animal.

5.0 Protección de datos

Acceso autorizado
Únicamente los integrantes del equipo listados arriba, profesores del curso y personal del CAETEC tienen acceso a estos datos.

Anonimización de datos
El dataset de vacas utilizado en este proyecto no requiere anonimización, ya que los identificadores contenidos (ID de vaca, métricas de salud y producción) no constituyen datos personales ni sensibles conforme a la Ley Federal de Protección de Datos Personales en Posesión de Particulares. Estos registros corresponden a animales de producción y no permiten rastrear información vinculada a personas físicas o productos específicos.

Normativa actual de la industria
En la industria ganadera y de agricultura de precisión, la gestión de datos se rige principalmente por marcos de trazabilidad y confidencialidad técnica más que por normativas de privacidad personal, ya que los registros se refieren a animales y operaciones productivas, no a personas físicas. En México, el marco aplicable es el establecido por el SINIIGA/SENASICA, que regula la identificación y trazabilidad pecuaria, mientras que a nivel internacional destacan la norma ISO 22005 (trazabilidad en la cadena alimento-pienso) y las directrices del ICAR (Animal Data Exchange) para intercambio seguro de datos entre sistemas.

En conjunto, estos lineamientos aseguran la confidencialidad y seguridad de la información sin requerir procesos de anonimización, dado que los identificadores del hato (ID de vaca, métricas de salud o producción) no son datos personales y su tratamiento se rige por criterios de trazabilidad técnica y confidencialidad comercial.

Almacenamiento seguro
Usaremos servicios de Amazon Web Services (AWS) para guardar nuestros datos de forma segura:

Amazon S3:
- Los datos están encriptados automáticamente

- Solo nuestro equipo tiene acceso mediante credenciales únicas
- Configuramos permisos específicos para cada tipo de archivo

Control de acceso:
- Cada miembro del equipo tiene su propio usuario con permisos limitados
- Usamos autenticación de dos factores (2FA)
- Los accesos quedan registrados automáticamente
- Configuramos políticas que sólo permiten acceso desde redes específicas

Configuración de seguridad:
- Bucket privado
- Encriptación en reposo activada
- Versionado de archivos para recuperación
- Logs de acceso habilitados
- Políticas de IAM restrictivas.

Tiempo de conservación de los datos
Durante el semestre mantendremos todos los datos necesarios para completar el trabajo. Tras la finalización del semestre, eliminaremos los datos originales pero se conservará el modelo entrenado.

6.0 Acceso y registro de manejo de datos

El sistema de control de acceso y registro adoptado por el proyecto se encuentra alineado con los principios y controles establecidos por la norma **ISO/IEC 27001**, estándar internacional para la gestión de la seguridad de la información.
En particular, el esquema implementado atiende los controles relacionados con:

- **A.5 – Políticas de seguridad de la información:** se establecen lineamientos explícitos para el acceso, manipulación y resguardo de datos.

- **A.6 – Organización de la seguridad:** el uso de roles IAM y llaves controladas garantiza la asignación de responsabilidades y el principio de privilegios mínimos.

- **A.9 – Control de acceso:** se utiliza autenticación obligatoria, 2FA y permisos estrictamente necesarios para lectura y escritura, evitando accesos no autorizados.

- **A.12 – Operación segura:** el sistema de logging registra cada operación de tratamiento de datos, permitiendo trazabilidad completa, detección de anomalías y capacidad de auditoría operacional.

- **A.16 – Gestión de incidentes:** la existencia de logs unificados, estandarizados y replicados localmente permite identificar eventos críticos y responder oportunamente ante incidentes de seguridad.

- **A.17 – Continuidad del negocio:** el modelo de redundancia —almacenamiento simultáneo en AWS S3 y respaldo local— contribuye a la disponibilidad y resiliencia ante fallos del servicio o interrupciones inesperadas.

El diseño del sistema no solo cumple con buenas prácticas de ingeniería de datos, sino que incorpora explícitamente los principios de **confidencialidad, integridad y disponibilidad (CIA)** requeridos por ISO/IEC 27001.

Con ello, el proyecto garantiza que todo acceso, transformación, lectura o escritura de datos quede documentado, autenticado y trazable, asegurando una gestión conforme a estándares internacionales de seguridad de la información.

Previo a la implementación del sistema automatizado, el control de uso de modelos y procesamiento de datos se realizaba mediante bitácora manual interna por medio de un canal de Whatsapp.
El registro histórico correspondiente consta del siguientes únicos registros:

Fecha	**Evento documentado**	**Hora registrada**
12/10/2025	Descarga dataset datos csv fuente original	17:30
13/10/2025	Transformación de datos, merge datos vacas	11:05
25/11/2025	Ejecución Modelo Sanidad · Isolation Forest v2.0	10:39
25/11/2025	Entrenamiento Modelo Comportamiento · Baseline/MLP	12:25
25/11/2025	Ejecución ETL v2.0 · consolidación dataset	12:44

Este esquema operó temporalmente como evidencia de control, pero no satisfacía las exigencias de auditoría, escalabilidad y seguimiento automatizado exigidas para una política de gestión de datos completa.

Sistema formal de registro (a partir del 25 de noviembre de 2025)

Desde dicha fecha quedó adoptado el nuevo esquema de logging unificado basado en **Amazon Web Services**, reemplazando el registro manual y permitiendo auditorías más precisas y automáticas.

Cada operación relacionada con datos genera automáticamente un registro que contiene visuales) quedan almacenadas en:

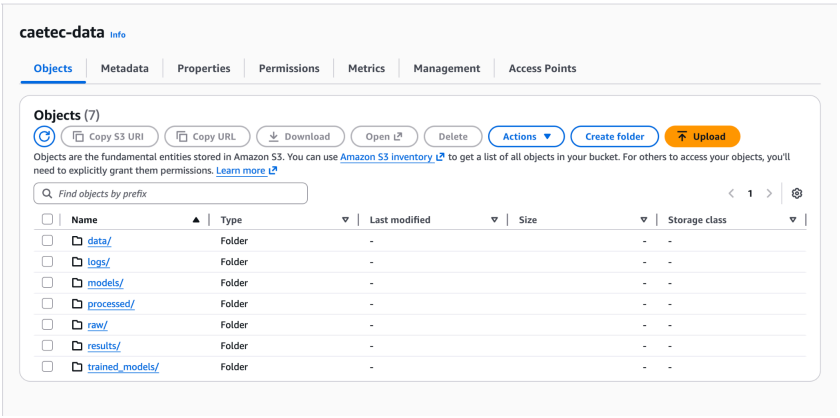
Campo	**Descripción**
date	Fecha exacta del evento
time	Timestamp preciso HH:MM:SS

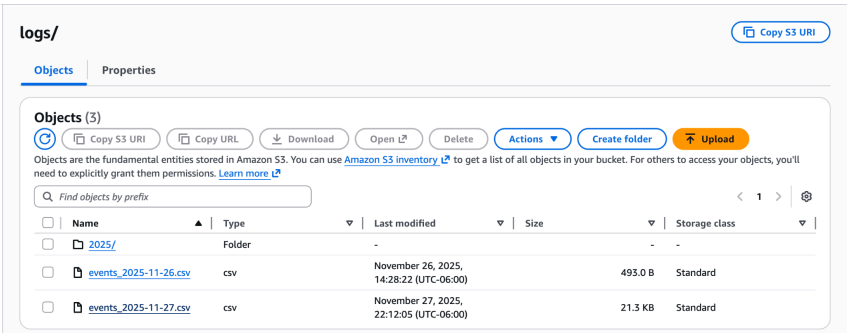
action	Acción realizada (read/write/delete/etc.)
resource_type	Tipo de recurso procesado (results, models, data, logs)
file	Objeto generado o consumido
script	Proceso o módulo responsable
extra	JSON enriquecido con contexto (shape, parámetros, objetivo del run, etc.)

Ejemplo registro real:
2025-11-27,10:38:09,read,data,data/sessions_health.csv,sanidad_iso_v2.py,"{""purpose"": ""iso_sanidad_train_v2"", ""source"": ""s3"", ""shape"": [7239, 20]}"

Ruta de logs:

Amazon S3 → caetec-data/logs/events_YYYY-MM-DD.csv





Si se hace algún registro el mismo día sobreescribe el archivo correspondiente al día, agregando los logs, si es un día distinto crea el nuevo archivo correspondiente.

Además nuestro proyecto cuenta con un sistema de redundancia, ya que no solo guarda los registros en el servicio de AWS sino que también se guardan localmente en la computadora de quien lo ejecute para así tener un respaldo en dado caso de algún fallo imprevisto con la conexión o el servicio de AWS.

Acceso y permisos de lectura escritura

El acceso a estos registros se encuentra restringido mediante IAM, con privilegios mínimos asignados por rol y autenticación obligatoria mediante 2FA. Ningún movimiento de datos es posible sin generación de log. Solo los integrantes de Concentrados que firmen la declaración de confidencialidad en el presente documento, tienen acceso a la llave IAM que permite la operación y acceso de datos, llave la cual será proporcionada junto variables de entorno de manera privada por el integrante Jesus Ramirez Delgado quien tiene usuario *root* del servicio de AWS.

El sistema de control de acceso y registro implementado garantiza trazabilidad documental y evidencia técnica de toda manipulación de datos, permitiendo auditorías, control de uso y respuesta oportuna ante incidentes.

Estado operativo actual

El sistema está en operación y absolutamente todos los modelos y archivos que ejecuten una lectura o escritura de datos cuenta con *endpoints* que permiten las operaciones. A la fecha de hoy 1 de diciembre de 2025, se han realizado operaciones:

Registro	Descargar ejemplo de registro
events_2025-11-26	click aquí
events_2025-11-26	click aquí

7.0 Bases legales

Esta política está fundamentada bajo la Ley Federal de Protección de Datos Personales en Posesión de Particulares de México, especialmente lo establecido en los artículos 9, 60 y 61:

- Artículo 9: Los responsables del tratamiento de datos deben cumplir con 9 principios rectores para la protección de los datos
    - Licitud
        - El responsable está obligado a hacer el tratamiento con apego y cumplimiento a lo dispuesto por la legislación mexicana y el derecho internacional
    - Consentimiento
        - El responsable deberá obtener el consentimiento para el tratamiento de los datos personales, a menos que no sea exigible
    - Información

        - El responsable deberá dar a conocer al titular la información relativa a la existencia y características principales del tratamiento a que serán sometidos sus datos personales a través del aviso de privacidad
    - Calidad
        - Se cumple con el principio de calidad cuando los datos personales tratados sean exactos, completos, pertinentes, correctos y actualizados según se requiera para el cumplimiento de la finalidad para la cual son tratados
    - Finalidad
        - Los datos personales sólo podrán ser tratados para el cumplimiento de la finalidad o finalidades establecidas en el aviso de privacidad
    - Lealtad
        - Establece la obligación de tratar los datos personales privilegiando la protección de los intereses del titular y la expectativa razonable de privacidad,
    - Proporcionalidad
        - Sólo podrán ser objeto de tratamiento los datos personales que resulten necesarios, adecuados y relevantes en relación con las finalidades para las que se hayan obtenido
    - Responsabilidad
        - El responsable deberá adoptar medidas para garantizar el debido tratamiento, privilegiando los intereses del titular y la expectativa razonable de privacidad.

- Artículo 60: El responsable deberá determinar las medidas de seguridad aplicables a los datos que maneje de acuerdo a 4 factores
    - El riesgo inherente por tipo de dato personal
    - La sensibilidad de los datos personales tratados
    - El desarrollo tecnológico, y
    - Las posibles consecuencias de una vulneración para los titulares
    - Tomar en cuenta los siguientes aspectos
        - El número de titulares
        - Las vulnerabilidades previas ocurridas en los sistemas de tratamiento
        - El riesgo por el valor potencial cuantitativo o cualitativo que pudieran tener los datos personales tratados para una tercera persona no autorizada para su posesión, y
        - Demás factores que puedan incidir en el nivel de riesgo o que resulten de otras leyes o regulación aplicable al responsable

- Artículo 61: El responsable debe realizar 9 acciones para garantizar la seguridad de los datos
    - Elaborar un inventario de datos personales y de los sistemas de tratamiento
    - Determinar las funciones y obligaciones de las personas que traten datos personales
    - Contar con un análisis de riesgos de datos personales que consiste en identificar peligros y estimar los riesgos a los datos personales
    - Establecer las medidas de seguridad aplicables a los datos personales e identificar aquéllas implementadas de manera efectiva
    - Realizar el análisis de brecha que consiste en la diferencia de las medidas de seguridad existentes y aquéllas faltantes que resultan necesarias para la protección de los datos personales

- Elaborar un plan de trabajo para la implementación de las medidas de seguridad faltantes, derivadas del análisis de brecha
- Llevar a cabo revisiones o auditorías
- Capacitar al personal que efectúe el tratamiento
- Realizar un registro de los medios de almacenamiento de datos personales

8.0 Incidentes y contingencias

Incidente	**Acción inmediata**
Pérdida de datos	Recuperar desde los respaldos versionados de los registros locales y documentar la causa de la pérdida de datos.
Divulgación accidental	Notificar al CAETEC inmediatamente, evaluar el impacto, aplicar las medidas correctivas correspondientes.
Falla de infraestructura	Utilizar respaldos de registros locales, documentar y evaluar la migración del sistema

DECLARACIÓN DE CONFORMIDAD

Los miembros del equipo declaran haber leído, entendido y aceptado cumplir con esta política de datos:

- Daniel Queijeiro Albo - A01710441
- Diego Alfaro Pinto - A01709971
- Diego Isaac Fuentes Juvera - A01705506
- Jesus Ramirez Delgado - A01274723
- Mauricio Anguiano Juarez - A01703337
- Luis Adrián Uribe Cruz - A01783129