

Privacidad y seguridad de los datos

Autora: Andrea Medina Rico

Matrícula: A01705541

Abstracto. Este reporte indaga en el proceso a seguir para asegurar la privacidad y seguridad de los datos proporcionados por el Socio Formador.

I. EVIDENCIA INDIVIDUAL

a. Estados de los datos

Sobre los datos de producción de leche, se recibió un archivo .csv por vaca en donde cada fila contiene información de una ordeña. Además, existen tres archivos con los identificadores de cada vaca enlistados. El primero es un inventario con características de las vacas. El segundo es un reporte de las patadas que han dado las vacas. El tercero es un reporte de sus estados en general. Fuera de esta información, no hay registros que se vinculen con datos personales. El archivo final utilizado para el entrenamiento del modelo es una combinación de los archivos individuales de las vacas, sin incluir el identificador de las mismas.

Las imágenes proporcionadas presentan tres escenarios distintos. Las de la carpeta MULTI son imágenes del corredor al que salen las vacas tras su ordeña. En la carpeta QUEUE, hay vacas en la zona de espera para ser ordeñadas. En las imágenes de BEDS, se observan las vacas en los echaderos. En algunas de las imágenes, resaltan los aretes de las vacas junto con su identificador. Para el entrenamiento, las imágenes fueron segmentadas (el resto es un fondo negro).

Ninguna de las fuentes de información presenta sus datos anonimizados completamente. Por parte del equipo, no fueron anonimizados por dos razones principales.

La primera es que mantener el identificador de las vacas es fundamental para el cumplimiento del objetivo del proyecto, el cual es "Generar conocimiento sobre la posibilidad de identificar a las vacas únicamente a través de sus patrones de manchas". Esto implica que el Socio Formador pueda subir una fotografía y el resultado del modelo regrese el identificador de esa vaca. En caso de no mostrarse ese *id*, el resultado no sería de valor para el cliente. En todo caso, podría tenerse un *id* genérico dentro del equipo para posteriormente vincularse al *id* real. Sin embargo, se decidió mantener esta información en el sistema de almacenamiento (Drive) ya que, de una u otra forma, los identificadores reales debían ser almacenados en alguna parte.

La segunda razón es la implementación de la solución. Se propone que el proceso de identificación de vacas se realice en el mismo entorno local donde se encuentran las imágenes: una Raspberry Pi. Tras la toma de la fotografía, la imagen se recibe en la Raspberry y, ahí mismo, se segmenta la imagen y se genera la predicción del modelo. La imagen no sale de este entorno local a la cual únicamente el SF tiene acceso. Por otra parte, se propone que la predicción de anomalías pueda

realizarse a través de una interfaz en donde el SF sube el documento de una vaca y visualiza el resultado. En este caso, el archivo no es almacenado. Únicamente regresa los datos necesarios.

b. Normativa actual

La NORMA OFICIAL MEXICANA NOM-001-SAG/GAN-2015 presenta los lineamientos para la identificación de las vacas. Aborda el tema de que la vaca debe contar con un identificador único y tenerlo presente en todo momento, mas no habla sobre la confidencialidad de estos datos.

II. EVIDENCIAS DE EQUIPO

En la política de acceso, se define la gestión de cambios, almacenamiento y acceso de los datos, así como las responsabilidades de los usuarios, para garantizar la seguridad de los mismos.

https://abalone-timbale-4a2.notion.site/Pol-tica-de-acceso-287a3ea0df618097b208f5f04539b31c?source=copy_link

En la bitácora de cambios, se registran los cambios realizados para asegurar la trazabilidad e integridad de los datos.

https://docs.google.com/spreadsheets/d/1H4hMijn9OsfHKDEhNZj7_t2ieoMNB2_zAa0QC1PRjc/edit?usp=sharing