



Tecnológico de Monterrey

Campus Querétaro

Preparación de los datos

Gamaliel Marines Olvera	A01708746
Uri Jared Gopar Morales	A01709413
José Antonio Miranda Baños	A01611795
María Fernanda Moreno Gómez	A01708653
Oskar Adolfo Villa López	A01275287
Luis Ángel Cruz García	A01736345

Inteligencia artificial avanzada para la ciencia de datos II
Grupo 501

Selección de imágenes

Las imágenes fueron obtenidas directamente de parte del Socio Formador, dentro de estas imágenes se contaban dos tomas, una de ellas con vista hacia la fila de ordeño y otra hacia las camas, siendo esta última la de valor para nosotros como equipo, resultando en un total de 9,634 imágenes, en las cuales a primera vista pudimos observar que capturan una sección con 3 camas completas y una incompleta, además de estar en un plano cenital. Todas las fotografías fueron tomadas desde la misma posición, sin embargo, es importante mencionar que se nos brindó la posibilidad por parte del Socio Formador de pedir más imágenes que fueran tomadas desde otro punto.

Limpieza de imágenes

De las imágenes originales, algunas contienen distintas características de luz, llegando a estar muy saturadas o muy oscuras. Sin embargo, se decidió conservar estas imágenes, con el objetivo de que el modelo pueda trabajar en este tipo de condiciones.

Se desecharon imágenes corruptas que se mostraban completamente negras, además de imágenes en las que no era posible distinguir la clase. Este descarte fue de 166 imágenes recortadas, y se realizó al momento de la clasificación. Al ser pocas, no impactaron de manera significativa el tamaño del dataset.

Construcción del dataset

Del dataset original (proporcionado por nuestro socio formador) que contenía 9,634 imágenes disponibles del área de camas de un único corral en el CAETEC, se utilizaron 3,997 imágenes para transformarlas:

1. De las 11 personas que conformamos los dos equipos de “Camas”, nos repartimos 7 personas 363 imágenes y 4 personas 364 para su posterior clasificación. Las imágenes fueron divididas y disponibles en carpetas personales en Google Drive.
2. Se preparó un script donde se trataron las imágenes para quitarles el ojo de pescado, con el fin de normalizar las imágenes y facilitar su procesamiento.
3. Se recortó a las imágenes los primeros 225 píxeles, pues estos píxeles correspondían a la mitad de una cama, lo que no nos iba a servir, pues necesitamos camas completas para su análisis.
4. Se dividió cada imagen en 3, con el fin de recortar cada una de las camas y así poder hacer el análisis de cada uno de los estados de las camas de manera más sencilla y práctica para el tiempo y de acuerdo a nuestro plan.
5. Una vez con las imágenes tratadas y recortadas, cada uno de nosotros, con las imágenes que nos tocaron, clasificamos cama por cama, el estado de

cada una de ellas. Esto se logró por medio del mismo código de Python, donde se ponía “1” para las camas con una vaca de pie, “2” para las camas con una vaca acostada y “3” para las camas vacías, con el fin de optimizar su clasificación. Una vez que todas las camas habían sido clasificadas, se guardaban en carpetas locales de cada uno de nosotros.

6. Para finalizar, se subió al mismo Drive las carpetas en formato zip (para no abarcar tanto espacio disponible de la nube) de cada una de las clasificaciones de las vacas de todos los miembros.
7. Se unieron todas las carpetas de cada clasificación de las camas de cada miembro en una sola, obteniendo 3 al final (una para cada estado) con el fin de unir todos los datos en el mismo Drive.

Atributos generados

Al clasificar las imágenes recortadas previamente, se clasificaron con el fin de etiquetar cada cama con 3 atributos posibles:

- Cama vacía
- Cama con vaca parada
- Cama con vaca acostada

Es decir, una imagen de la cama puede tener un único atributo de estos, que mide el estado de la cama.

Registros generados

De las 3,997 imágenes originales, que fueron divididas entre 3 (por las 3 camas) quedando un total de 11,991 imágenes, que después de la limpieza, quedaron 11,825 imágenes de camas individuales, de las cuales se generaron:

- 8,803 archivos de imagen de camas vacías
- 2,625 archivos de imagen de camas con vaca acostada
- 397 archivos de imagen de camas con vaca parada

Integración de las imágenes

Las imágenes divididas por cada miembro de los equipos se unificaron en una sola carpeta primaria, que a su vez, estaba dividida en tres carpetas con cada una de las clasificaciones de las camas, con el fin de juntar toda la información y facilitar el entrenamiento del modelo.

El dataset para el modelado, está disponible en el Google Drive acordado por el equipo.