



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Querétaro

## **Fase de Entendimiento de Negocio**

### **Reporte de Descripción de los Datos**

#### **Autores:**

A01368818 Joel Sánchez Olvera

A01661090 Juan Pablo Cabrera Quiroga

A01704076 Adrián Galván Díaz

A01708634 Carlos Eduardo Velasco Elenes

A01709522 Arturo Cristián Díaz López

**TC3007C.501**

Inteligencia artificial avanzada para la ciencia de datos II

#### **Fecha:**

9 de Octubre del 2024

## **Reporte de descripción de los datos**

### **Observaciones y Formato de los Datos**

Durante la fase de clasificación de las imágenes, se trabajó con un conjunto de datos compuesto por 9634 imágenes en formato JPG, con un tamaño promedio de entre 600 y 650 KB y una resolución de 1920x1080 píxeles. De estas, se seleccionaron aleatoriamente 4004 imágenes para el análisis. Estas imágenes fueron divididas entre los miembros del equipo, y cada uno clasificó 364 imágenes.

### **Observaciones sobre la Distribución de los Datos**

Al analizar las imágenes, el equipo observó que había una notable prevalencia de imágenes con camas vacías, en comparación con las imágenes que mostraban vacas paradas o acostadas. Esto es un aspecto importante, ya que indica que el comportamiento de las vacas al utilizar las camas es menor de lo esperado, o que los datos no capturaron suficiente actividad. Este desequilibrio en las clases deberá ser considerado al momento de entrenar los modelos de machine learning, para evitar un sesgo hacia la clase mayoritaria.