



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Querétaro

Fase de Entendimiento de Datos

Reporte de Exploración de 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 exploración de datos

Primeros Descubrimientos

Durante la fase inicial de exploración de datos, hemos realizado varios descubrimientos significativos. Todos los descubrimientos que encontramos, son derivados de una colaboración en conjunto entre los dos equipos que elegimos trabajar con las imágenes de las camas.

Cabe destacar que para hacer esto, utilizamos un script para etiquetar las imágenes, este script las recorta en tres partes, correspondientes a la cantidad de camas en cada imagen.

Estas son nuestras conclusiones:

- Notamos que una gran cantidad de ellas muestran camas vacías (8,160) en comparación con la cantidad de fotos de vacas acostadas (2,338) y de vacas paradas (359).
- Identificamos que ciertas imágenes están completamente oscuras o excesivamente brillantes. Asociamos esto con los momentos en el día en el que se tomaron las fotografías (de noche y al amanecer).
- Las imágenes resultantes son de tamaño 450 x 950

Hipótesis Inicial y su Impacto en el Resto del Proyecto

Nuestra hipótesis inicial es que existen patrones significativos en el uso de las camas por parte de las vacas que pueden ser identificados y analizados mediante

técnicas de procesamiento de imágenes y minería de datos. Específicamente, suponemos que:

- Las vacas utilizan ciertas camas más que otras (específicamente, la cama 3 no se utiliza casi nada en comparación con las otras dos camas. Esto puede ser que la cama 3 tiene una pared a la derecha, que es donde se encuentra el bebedero), indicando preferencias o comportamientos que afectan la eficiencia del espacio de descanso.
- Es posible cuantificar el tiempo que las vacas pasan en las camas, lo que nos permitirá calcular el tiempo promedio de uso y detectar posibles ineficiencias.
- Podemos identificar y clasificar con precisión la presencia y posición de las vacas (paradas o acostadas) en las imágenes, a pesar de las variaciones en iluminación y calidad.

El impacto de esta hipótesis en el proyecto es que, al confirmar estos patrones, podremos optimizar la distribución y uso de las camas, responder a las preguntas clave de negocio y cumplir con los criterios de éxito establecidos, como alcanzar una precisión de al menos 80% en la clasificación y mantener un margen de error bajo. Esto facilitará que el personal de CAETEC pueda implementar mejoras basadas en hallazgos significativos respecto al uso de las camas.

Gráficas de la distribución de los datos.

