



Instituto Tecnológico y de Estudios Superiores de  
Monterrey Campus Querétaro

## **ETAPA DE EVALUACIÓN**

### **Autores:**

A01368818 Joel Sánchez Olvera

A01661090 Juan Pablo Cabrera Quiroga

A01704076 Adrián Galván Díaz

A01708634 Carlos Eduardo Velasco Elenes

A01709522 Arturo Cristián Díaz López

### **Fecha:**

22 de noviembre del 2024

## Introducción

La etapa de evaluación en la metodología CRISP-DM es crucial para validar si los modelos desarrollados cumplen con los objetivos establecidos, tanto a nivel técnico como de negocio. Durante esta fase, se analizan los resultados obtenidos a partir de los modelos implementados, evaluando su desempeño con base en las métricas clave, y se examina su alineación con los objetivos de minería de datos y de negocio. Además, esta etapa busca identificar áreas de mejora y proponer ajustes necesarios para optimizar el sistema.

El objetivo principal de esta evaluación es garantizar que los resultados generados no solo sean precisos y confiables, sino que también proporcionen valor en términos de análisis de datos y toma de decisiones. En el contexto del proyecto CAETEC, se busca verificar que el sistema desarrollado permita cuantificar y caracterizar la población ganadera en función de su postura y ubicación, además de facilitar la integración de estos resultados en los flujos de trabajo operativos a través de bases de datos accesibles y formatos estándar.

## Evaluación de los Resultados de Minería de Datos

Los resultados obtenidos están alineados con los objetivos del proyecto, destacando un desempeño sobresaliente en la identificación y clasificación de vacas según su postura (parada/acostada) y en el análisis de ocupación de espacios. Cada modelo desarrollado cumplió con los criterios de éxito establecidos y superó los umbrales mínimos de precisión previamente establecidos. Los modelos generaron registros precisos de las posiciones y actividades de las vacas, permitiendo un análisis del uso de las áreas de descanso. A continuación se listan los desempeños de los diferentes modelos:

### Modelo de Clasificación de Posturas

Precisión: 96%

Recall: 96%

F1-Score: 96%

Curva ROC: AUC = 0.97

Este modelo, basado en una CNN implementada en TensorFlow, manejó exitosamente un dataset desbalanceado utilizando técnicas de data augmentation y ajuste de pesos por clase. Logró un balance adecuado entre precisión y recall, clasificando con eficacia las posturas de las vacas en imágenes aéreas.

### Modelo de Detección de Vacas

Precisión: 98.7%

Recall: 98.1%

mAP@0.5: 97.5%

mAP@0.5:0.95: 80%

Utilizando YOLOv5, este modelo alcanzó una excelente generalización en la detección de vacas en imágenes complejas y condiciones de iluminación variables. Su capacidad de identificar bounding boxes con alta precisión contribuyó directamente al conteo y caracterización del ganado en las áreas de descanso.

### Análisis de Patrones en Camas de Arena

Métricas GLCM: Contraste ( $\approx 15-20$ ), Homogeneidad ( $< 1.0$ ), Correlación baja.

Periodicidad: Ausencia de patrones repetitivos.

Detección de Bordes: Consistente en diferentes métodos (Sobel, Canny, Laplaciano).

Este análisis reveló una ocupación uniforme del espacio, sin patrones de sobreuso o compactación excesiva. Los resultados respaldan la óptima distribución del ganado en las áreas disponibles.

El desempeño individual de los modelos desarrollados está directamente alineado con los objetivos de minería de datos. El modelo de detección de vacas permite identificar y contar con precisión las bounding boxes, generando registros detallados en la base de datos, mientras que el modelo de clasificación caracteriza la postura de las vacas en cada área monitoreada, enriqueciendo estos registros con información clave como su ubicación y precisión. Por su parte, el análisis de patrones en camas de arena complementa estos resultados al evaluar la ocupación y uso uniforme del espacio. A continuación se listan los puntos claves del cumplimiento de los objetivos de minería de datos::

### **Conteo de bounding boxes**

El sistema identifica y contabiliza las bounding boxes correspondientes a las vacas detectadas en las imágenes aéreas, registrando estos resultados en la Tabla 1. Esta tabla incluye variables como el identificador de la imagen (id\_img\_pk), la marca de tiempo (timestamp) y el número total de vacas detectadas (cow\_count), cumpliendo con los criterios de éxito definidos.

### **Clasificación de vacas por postura**

El sistema clasifica cada vaca en una de las dos categorías posibles: "parada" o "acostada". Los resultados se almacenan en la Tabla 2, que incluye variables como el identificador único de la vaca detectada (id\_detected\_cow\_pk), su ubicación (cow\_centroid), y la precisión del modelo (prediction\_accuracy). Esta estructura permite analizar de manera exhaustiva la postura y distribución del ganado en las áreas monitoreadas.

## **Integración con base de datos y exportación**

El sistema permite una conexión eficiente a la base de datos a través de SSH o conexiones locales, proporcionando capacidades de exportación en formatos comunes como CSV y XLSX. Esto garantiza la accesibilidad y usabilidad de los resultados, facilitando su análisis y aprovechamiento por parte de los responsables del proyecto.

En conjunto, estos logros permiten que el modelo de detección y clasificación no solo cuantifique y caracterice la población ganadera, sino que también brinde una base sólida para la toma de decisiones estratégicas relacionadas con la ocupación de espacios y el bienestar animal.

[Evaluar si hay nuevos objetivos de negocio que deben abordarse después o si deben crearse nuevos proyectos]

Establecer conclusiones para futuros proyectos de minería de datos]

## **Modelos aprobados**

Como se demostró en la sección anterior, los modelos desarrollados cumplen satisfactoriamente con los criterios de éxito establecidos tanto a nivel técnico como de negocio. Cada modelo logró alcanzar las métricas definidas, garantizando la precisión y confiabilidad necesarias para la detección, clasificación y análisis de patrones. Además, los resultados generados permiten satisfacer los objetivos de minería de datos, proporcionando información clave sobre la ocupación y uso de las áreas de descanso; el conjunto de modelos construidos para la solución final permiten que se determine la ocupación de los espacios destinados al descanso de las vacas en el rancho, identificando la población de ganado en cada zona y evaluando qué áreas están ocupando la mayor parte del tiempo.

## **Revisión del proceso**

El proceso de minería de datos siguió todas las etapas de CRISP-DM de manera estructurada, garantizando un flujo claro desde la comprensión del negocio hasta la implementación de los modelos. Sin embargo, se identificó que la etapa de modelado tomó más tiempo del recomendado debido a la necesidad de realizar iteraciones exhaustivas para ajustar cada modelo. Este enfoque, aunque prolongado, aseguró la robustez y confiabilidad de los resultados obtenidos.

Durante el desarrollo de los modelos, se encontraron fallas en las primeras iteraciones o versiones, las cuales fueron solucionadas a través de ajustes documentados (hacer

referencia a la tabla correspondiente). No se presentaron resultados engañosos o inesperados, ya que estos fueron corregidos durante el proceso iterativo.

Un área de mejora clave fue la forma en que se manejaron las iteraciones. En lugar de realizar múltiples ciclos cortos de modelado como recomienda CRISP-DM, se realizó una única iteración prolongada que tomó aproximadamente seis semanas. Aunque este enfoque funcionó en este proyecto, formalizar las iteraciones en ciclos más cortos hubiera permitido una mayor adaptabilidad y velocidad en la corrección de errores, lo que constituye una oportunidad para futuros proyectos.

Gracias a este proceso, se logró determinar la ocupación de los espacios destinados al descanso de las vacas, identificando la población en cada zona y evaluando las áreas con mayor ocupación.

Criterios de éxito:

- Identificar el número total de vacas en cada área específica (conteo de población).
- Mapear la distribución de las vacas en el espacio geográfico asignado para evaluar la ocupación.
- Generar registros accesibles en formatos estándar (CSV y XLSX) para facilitar la visualización y el análisis.

## **Lista de posibles acciones**

Una posible acción futura es integrar los análisis de patrones de arena con los criterios de éxito de minería de datos. Esto permitiría generar insights más valiosos al analizar no solo la ocupación de los espacios, sino también cómo las texturas y patrones de uso impactan la calidad de los descansos y el bienestar de los animales.

Los recursos actuales son suficientes para implementar esta acción. Sin embargo, el desafío principal no radica en los recursos materiales o técnicos, sino en gestionar mejor el tiempo asignado para completar el proyecto dentro de los plazos establecidos.

Se recomienda formalizar las iteraciones en proyectos futuros, dividiendo las fases largas como el modelado en ciclos más pequeños y manejables. Esto permitirá:

- Mayor flexibilidad para realizar ajustes.
- Identificar problemas más rápidamente y evitar tiempos prolongados en una sola fase.
- Mantener un mejor control sobre el cronograma del proyecto.

## **Decisión**

Después de analizar las posibles acciones, se decidió no llevar a cabo la integración de los análisis de patrones de arena con los resultados de los modelos de minería de datos. Esta decisión se tomó considerando el tiempo limitado restante para completar el proyecto.

En lugar de realizar la integración propuesta, se optó por enfocar los esfuerzos en proporcionar insights de valor directamente alineados con las necesidades del socio formador. Esta decisión permite comprender de manera más eficiente las prioridades del negocio, entregando resultados claros, accionables y en un formato accesible que pueda ser implementado de inmediato. Este enfoque asegura que los objetivos del proyecto se cumplan dentro de los plazos establecidos, maximizando el impacto para el socio.