



**Campus Querétaro**

Evaluación

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

## Introducción

Este documento presenta la fase de evaluación del modelo desarrollado, siguiendo el marco metodológico CRISP-DM (Cross-Industry Standard Process for Data Mining). En esta etapa, se analiza si el modelo cumple con el objetivo de negocio planteado inicialmente y si los hallazgos obtenidos son consistentes con las metas definidas, además se evalúan los resultados obtenidos a lo largo del proyecto y si estos son de valor o no para que sean entregados al socio.

El principal objetivo de negocio de este proyecto es identificar patrones en el estado de ocupación de las camas en tiempo real, permitiendo un análisis continuo y automatizado de las imágenes recibidas. Para lograr nuestro objetivo, realizamos un modelo de redes neuronales convolucionales (CNN) para la clasificación de los estados de ocupación de las camas, obteniendo así este estado de ocupación en tiempo real, es decir, cada que llega una imagen al dataset.

En las iteraciones de modelado realizadas, se alcanzaron mejoras significativas, superando los criterios de éxito establecidos. El modelo final logró una precisión del 96% en el conjunto de prueba, lo que demuestra su capacidad para identificar los estados de las camas con un alto grado de confiabilidad. Asimismo, los resultados de las predicciones mostraron un rendimiento consistente durante diferentes horarios del día.

A continuación, se describen los avances logrados en cada iteración, junto con un análisis detallado del proceso seguido, los desafíos encontrados y las oportunidades de mejora identificadas. Este análisis tiene como objetivo brindar un panorama integral del impacto del proyecto y sentar las bases para futuros desarrollos relacionados.

## Evaluación de resultados de minería

Como recapitulación, cabe destacar que el objetivo de negocio que se busca cumplir es el siguiente:

- Se obtiene conocimiento de los estados de las camas (vacía/vaca acostada/vaca parada) en tiempo real, es decir, cada que llegue una imagen al dataset.

Este objetivo puede ser evaluado por medio de las métricas obtenidas en un ambiente de producción simulado y posteriormente validado por el socio formador.

El criterio de éxito básicamente es saber si se obtiene conocimiento de los estados de las camas (vacía/vaca acostada/vaca parada) en tiempo real, es decir, cada que llegue una imagen al dataset.

Por otra parte, el objetivo de minería de datos que se logró durante el modelado es el siguiente:

- Entregar información del estado de las vacas en las camas (vacía/acostada/parada), junto con la etiqueta del tiempo para cada una de las fotos que se reciben.

Este siendo evaluado por medio de dos criterios de éxito:

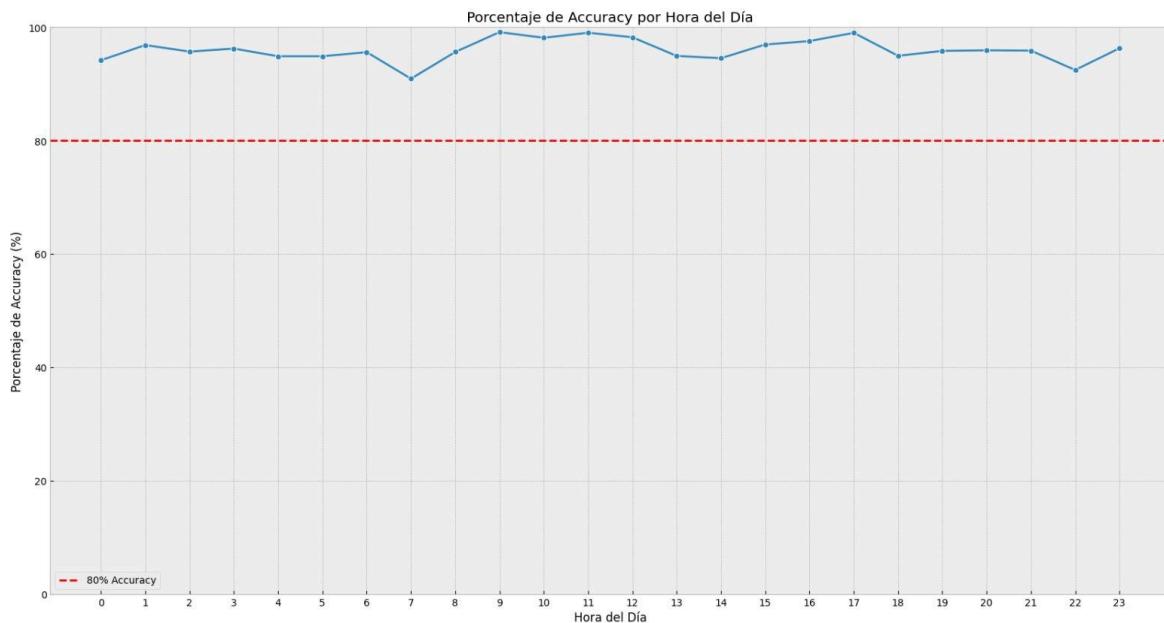
- El modelo identifica con un 80% de precisión el estado de cada una de las camas.
- El modelo entrega el análisis de todas las camas en el 80% de las imágenes a lo largo del día.

El resultado obtenido en la última iteración de modelado indica que se cumplen los criterios establecidos, ya que la precisión del modelo en el conjunto de prueba es del 96%, siendo significativamente mayor al porcentaje propuesto en los criterios de éxito. Mientras que en las gráficas de predicciones correctas contra tiempo se observa que se cumple el segundo criterio, manteniendo predicciones acertadas en diferentes horas del día.

A continuación se muestra una comparación de los resultados obtenidos en cada iteración:

Aspecto	Modelo 1	Modelo 2	Modelo 3
Transformaciones en imágenes	Redimensionamiento, giros horizontales y verticales, rotación aleatoria.	Transformaciones del Modelo 1 + Zoom out	Transformaciones del Modelo 2 + Saturación (ColorJitter, factor de saturación = 2).
Épocas de entrenamiento	5 épocas	50 épocas	50 épocas
Pesos en función de pérdida	No se utilizaron pesos específicos para las clases.	Pesos ajustados: 4.0 para "vaca acostada" y 8.0 para "vaca de pie" debido al desbalance de clases.	Pesos ajustados: 3.0 para "cama vacía", 1.0 para "vaca acostada" y 9.0 para "vaca de pie".
Learning rate	0.001	0.001	0.0001
Desempeño en la clase "cama"	True Positives: 1300 False Positives: 44	True Positives: 1306 False Positives: 34	True Positives: 1312 False Positives: 40

vacía"	False Negatives: 23	False Negatives: 17	False Negatives: 11
Desempeño en la clase "vaca acostada"	True Positives: 375 False Positives: 52 False Negatives: 19	True Positives: 371 False Positives: 16 False Negatives: 23	True Positives: 356 False Positives: 13 False Negatives: 38
Desempeño en la clase "vaca de pie"	True Positives: 5 False Positives: 1 False Negatives: 55	True Positives: 30 False Positives: 20 False Negatives: 30	True Positives: 31 False Positives: 25 False Negatives: 29
Matriz de confusión	Mayor precisión en clases mayoritarias ("cama vacía"), pero menor en clases minoritarias.	Mejor distribución, aunque "vaca de pie" sigue siendo la clase más difícil de clasificar.	Distribución más robusta, aunque las clases minoritarias aún enfrentan dificultades.
F1-score (Test)	94%	98%	96%
Áreas de mejora	Manejo del desbalance de datos, aumento de transformaciones y mayor número de épocas de entrenamiento.	Mejorar la detección de la clase "vaca de pie".	Incrementar el dataset para clases minoritarias y mejorar aún más las estrategias de balance.
Accuracy en Test	95%	96%	96%
Accuracy promedio por cama	59.60%	80%	95%



**Imagen 1:** Accuracy promedio a lo largo del día.

Para confirmar que nuestro modelo es funcional durante el 80% del día (19.2 h) graficamos su desempeño promedio a lo largo del día, aunque el desempeño baja alrededor de las 7 de la mañana en general se mantiene arriba del 90% lo cual cumple con nuestra promesa de superar el 80% a lo largo del 80% del día.

Al cumplir con el objetivo de minería de datos, logramos un impacto significativo en los hallazgos respecto al descanso de las vacas, ya que podemos brindar conocimiento nuevo y útil al Socio Formador que antes no podía identificar de una manera precisa, eficiente y automatizada. Dentro de las aplicaciones que tiene este conocimiento es identificar si hay camas que deben ser revisadas debido a que las vacas no la utilizan, o saber si las vacas tienen alguna preferencia en su descanso a lo largo del día.

Por otra parte, en la prueba de arquitectura realizada, pudimos comprobar que el modelo tarda 7 segundos en realizar una predicción, lo que, sumado al cumplimiento del objetivo de minería de datos, nos permite concluir que el proyecto cumple con el objetivo de negocio inicial. Este tiempo de predicción resulta altamente favorable considerando el contexto de uso continuo del modelo. Cada ejecución implica un tiempo de carga inicial de 30 segundos y posteriormente 7 segundos por predicción. Esto significa que en un día típico, si el socio desea ejecutar este proceso cada 5 minutos (288 veces al día), el tiempo total invertido será de **57 minutos y 36 segundos por día** (calculado como  $288 \times (30 \text{ segundos de carga} + 7 \text{ segundos de predicción})/60$ ).

Comparado con tiempos más elevados que podrían resultar de arquitecturas menos optimizadas, este modelo proporciona una solución eficiente y práctica para realizar análisis en tiempo real. Además, la automatización que ofrece permite al socio formador acceder a resultados confiables sin incurrir en tiempos de espera significativos. Esto refuerza el valor del modelo no solo en términos de precisión y consistencia en sus predicciones, sino también en términos de practicidad y escalabilidad para un uso diario intensivo.

Pero, ¿Por qué son buenos estos tiempos?:

- **Minimiza el impacto en la operación diaria:** Con un tiempo total menor a una hora por día para análisis continuos, el socio puede integrar esta solución sin afectar su flujo de trabajo o comprometer tiempos críticos de decisión.
- **Eficiencia energética y de recursos:** Al reducir tiempos de ejecución y predicción, se optimiza el uso de recursos computacionales, lo que es clave para implementaciones en sistemas locales o en la nube.
- **Escalabilidad en tiempo real:** La arquitectura permite una integración fluida, asegurando que incluso con altas frecuencias de uso, el sistema mantiene su desempeño y confiabilidad.
- **Viabilidad para futuras expansiones:** En caso de que el socio quiera escalar el uso del modelo (por ejemplo, agregar más cámaras o monitorear más camas), los tiempos actuales muestran que el sistema es capaz de manejar volúmenes crecientes de datos sin penalizaciones severas de rendimiento.

Durante el proceso no se identificaron nuevos objetivos de negocio, sin embargo, sí se identificó un posible nuevo proyecto cuyo objetivo pueda ser el que se identifiquen comportamientos de vacas en específico. Como conclusión para futuros proyectos de minería de datos, se recomienda identificar la mayor cantidad de métricas que sean de utilidad para evaluar los criterios de éxito, además de mantener un control de versiones, un estándar de commits<sup>1</sup> en GitHub y una documentación correcta a lo largo de las diferentes etapas para obtener un mejor resultado, en adición, el seguimiento adecuado de un marco de trabajo, nos permitió organizar de mejor manera las actividades necesarias y el tiempo destinado a cada una de ellas.

## Revisión de proceso

En cuanto al proceso seguido durante este proyecto, se utilizó como marco de trabajo CRISP-DM.

---

<sup>1</sup> **Commit:** Un commit en GitHub se refiere a la captura instantánea de los cambios en un momento específico en un proyecto. Al hacer un commit, se confirma un conjunto de cambios provisionales de forma permanente.

Durante la fase de entendimiento del negocio, se siguieron de manera correcta las actividades indicadas por CRISP-DM, cambiando únicamente que no se consideró la definición de costos y beneficios, ya que actualmente no es un proyecto con un presupuesto económico. El correcto seguimiento nos permitió terminar la fase en el tiempo planeado, formando un objetivo de negocio y un objetivo de minería de datos que fue aprobado por el Socio Formador y por los profesores.

Por otra parte, en la fase de entendimiento de datos, se agregó al contenido una evaluación de los datos del proyecto para saber si eran considerados como Big Data<sup>2</sup> y con base en eso, se describieron posibles herramientas y elementos que se pudieran implementar. Esto nos ayudó a decidir que no se trataba de un caso de Big Data y nos permitió tener un acercamiento de posibles tecnologías que podría implementar el Socio Formador en caso de querer hacer uso de cómputo en la nube o de querer escalar el proyecto. Las actividades de esta fase se realizaron sin inconvenientes, permitiéndonos terminar en el tiempo especificado y brindándonos el entendimiento necesario para poder pasar a la siguiente etapa de forma completa.

Ya con el entendimiento de datos pasamos a la preparación de los mismos, en la cual no se hicieron adaptaciones significativas dentro del proceso de CRISP-DM, solo es importante mencionar que el enfoque fue destinado a imágenes y no a datos estructurados, por esto mismo y porque las imágenes ya contaban con formato en el nombre, se omitió la actividad de formatear datos. Lo que cabe destacar durante esta fase es que al discutir sobre las imágenes con las que contábamos, decidimos no seguir con una de nuestras ideas que era implementar bounding box<sup>3</sup>, por lo que esta etapa terminó 3 días antes de lo planeado.

Al tener ya listo el dataset limpio y transformado, se procedió con la etapa de modelado. Para esta etapa, las modificaciones a CRISP-DM que realizamos fue el incluir una [Introducción General del Modelado](#) para la etapa de Diseño de Pruebas. Este documento se incluyó debido a que en esta etapa, realizamos 3 iteraciones en la fase de modelado (ya que, con las métricas definidas, los resultados evaluados no cumplían con nuestro objetivo de minería de datos), y en cada una de estas iteraciones, se iban añadiendo nuevas métricas de evaluación debido a los hallazgos encontrados en cada fase y la necesidad que se tuvo en cada una de ellas para evaluar distintos aspectos de su desempeño. Otra modificación que se hizo en esta etapa fue el hacer un análisis de las predicciones obtenidas en los

---

<sup>2</sup> **Big Data:** Es un término utilizado para describir a los grandes volúmenes de datos, que, debido a su tamaño, complejidad y velocidad de crecimiento, dificultan su captura, gestión, procesamiento o análisis mediante tecnologías y herramientas convencionales.

<sup>3</sup> **Bounding box:** son etiquetas de regiones rectangulares que se utilizan para tareas de visión por computadora (CV). En Machine Learning (ML), un objeto puede ser detectado de manera automática con bounding boxes.

Modelos 2 y 3, de modo que pudiéramos conocer los hallazgos obtenidos a través de nuestro modelo, pues es una de las cosas que queremos entregar al final del proyecto. Una de las partes de esta etapa que se omitió fue la de Parámetros Revisados y Ajustados, pues cuando el modelo actual no era suficiente, no describíamos como tal los datos numéricos de los parámetros a modificar, sino que en la fase de Evaluación del Modelo, se nombraban posibles parámetros que se pudieran modificar en el siguiente modelo para que tuviera un mejor rendimiento. Esta omisión no estuvo errónea, pues no sabemos a ciencia cierta los números exactos que debemos de cambiar para que tenga un mejor desempeño el modelo.

Como elementos a mejorar en el proceso, se debieron considerar más métricas desde la primera iteración de modelado, para de esta manera detectar con mayor precisión qué aspectos no estaban funcionando de manera correcta, las cuales son:

- Accuracy promedio por cama: Detecta si las predicciones son correctas por cada cama, ya que cuentan con condiciones diferentes, como la iluminación.
- Gráficas de predicciones por tiempo: Detecta si las predicciones son correctas en las diferentes horas del día.

Creemos que pudimos haber evitado el no usar todas estas métricas desde un inicio de la etapa si hubiéramos leído papers con un enfoque similar de clasificación en áreas establecidas, y así identificar que a veces la métrica de accuracy del modelo no siempre es la mejor porque puede dar falsos positivos.

### Lista de posibles acciones y decisiones

Después de analizar la funcionalidad y la aplicación desarrollada, hemos concluido que con nuestra implementación podemos tomar las siguientes acciones:

Resultado	Cumple con Obj Negocio	Cumple con Obj Minería Datos	Potencial de Entrega	Potencial de Mejora	Evaluación	Decisión
Modelo V1	✗	✗	✗	✗	El resultado no cumple con los objetivos de negocio y minería de datos, por lo cual es despreciado.	No se entrega.
Modelo	✓	✗	✗	✓	El resultado cumple	No se entrega.



o V2					con un porcentaje de los objetivos de negocio y minería de datos; hay áreas de mejora. Considerando el tiempo y la viabilidad, decidimos mejorarlo.	
Modelo V3	✓	✓	✓	✗	El modelo se puede mejorar; sin embargo, por los tiempos no es viable buscar una mejora.	Se entrega.
Información de camas	✓	✓	✓	✗	La información obtenida por medio de la utilización del modelo es buena, se puede mejorar; sin embargo, por los tiempos no es viable buscar una mejora.	Se entrega.

## Terminología

- **Desbalance de clases:** Cuando hay más datos de un tipo que de otro (por ejemplo, muchas fotos de camas vacías, pero pocas de vacas de pie), lo que puede dificultar que el modelo aprenda correctamente.
- **Precisión (accuracy):** Qué tan bien el modelo realiza su tarea, expresado como un porcentaje.
- **Matriz de confusión:** Una tabla que muestra cuántas veces el modelo acierta o se equivoca al clasificar cada tipo de estado (vacía, acostada, de pie).
- **F1-score:** Una medida que combina precisión y sensibilidad para evaluar el desempeño del modelo, especialmente cuando las clases están desbalanceadas.

## Referencias

AulaB. (n.d.). *Commit en Git: Qué es, cómo funciona y por qué es tan importante*. Recuperado el 22 de noviembre de 2024, de <https://aulab.es/articulos-guias-avanzadas/62/commit-en-git#:~:text=Un%20commit%20in%20Git%20representa,repository%20en%20un%20momento%20espec%C3%ADfico>

PowerData. (n.d.). *Big Data*. Recuperado el 22 de noviembre de 2024, de <https://www.powerdata.es/big-data>