Deployment

Adrián Matute, Pablo Martínez, Osvaldo Del Valle, Andres Callealta, Jorge Martínez

*Tecnológico de Monterrey Campus Querétaro, México*

**Plan de entrega**

Opciones para la implantación del modelo híbrido (un solo modelo)

* Instalación por parte del equipo
  + Se agendará una visita al CAETEC donde el equipo de trabajo realizará la instalación del sistema en el entorno de producción. Posteriormente se dará una capacitación al personal y usuarios finales para el uso del sistema.
* Instalación por parte del cliente
  + Se entregará un repositorio de github con los modelos y el sistema de gestión. Estos incluirán:
    - Manual de instalación: para la que el sistema pueda ser instalado por parte del cliente.
    - Manual de usuario: Describe el funcionamiento y operación del sistema para que los usuarios finales puedan hacer uso de este.
    - Manual técnico: Descripción detallada de todos los componentes del sistema para su mantenimiento y/o trabajo futuro.

De las dos opciones consideradas para la entrega del modelo, aplicaremos la segunda (Instalación por el cliente). Esto vendrá con su respectivo manual y scripts para poder montar el sistema automáticamente.

**Plan de monitoreo y de mantenimiento**

El objetivo es asegurar que el sistema de detección de vacas y monitoreo de aglomeraciones continúe siendo preciso y relevante para optimizar la producción de leche en CAETEC.

Los encargados de operar el sistema considerarán los siguientes puntos y en caso de que no se cumplan se podrán tomar una serie de decisiones:

***Estrategia de monitoreo***

1. Definir métricas clave para monitoreo continuo:
   1. Precisión de modelo en condiciones de día y noche.
   2. Frecuencia de detecciones incorrectas (falsos positivos/negativos).
   3. Tiempos promedios de respuesta para identificar aglomeraciones.
   4. Impacto en la optimización de las filas de ordeño.
2. Establecer un sistema de alertas:
   1. Alertas automáticas cuando la precisión del modelo caiga por debajo del umbral de 85%.
   2. Notificaciones si las aglomeraciones persisten durante períodos grandes de tiempo.
3. Implementar pruebas periódicas:
   1. Comparación de datos reales de conteo contra las predicciones del modelo.

***Mantenimiento del sistema***

1. Actualización del modelo:
   1. Añadir imágenes de las diferentes estaciones del año para garantizar adaptabilidad
2. Revisión de hardware y software:
   1. Verificar que las cámaras se encuentren en buenas condiciones.
3. Documentación:
   1. Mantener un registro de todas las actualizaciones, problemas y soluciones implementadas.
   2. Generar reportes periódicos con recomendaciones de acuerdo a los datos recolectados.

***Criterios para dejar de usar el sistema***

1. La precisión del modelo cae de forma constante por debajo del umbral definido y no es recuperable con reentrenamientos.
2. El sistema no logra identificar de manera correcta aglomeraciones de vacas en diferentes momentos del día.
3. Cambios en los objetivos del negocio o en la estructura del proceso de ordeño.

***Acciones ante fallas***

1. Si el sistema no puede seguir usándose:
   1. Revisar los datos y realizar un reentrenamiento.
   2. Proponer nuevas arquitecturas de modelos.
   3. En caso de tener fallas críticas, iniciar un nuevo proyecto enfocado en cubrir los aspectos con deficiencias.
2. Si los datos recolectados ya no reflejan condiciones actuales:
   1. Realizar una nueva recolección de datos para ajustar el modelo.

Dado que la meta final del proyecto, es ayudar al CAETEC a mejorar la producción de leche, es importante hacer una revisión constante del sistema para saber si los datos generados ayudan a evitar aglomeraciones. Además, es importante determinar si los beneficios económicos provenientes del sistema justifican que se le dé un mantenimiento de manera continua.

**Reporte de producto final**

**Resultados obtenidos:**

El proyecto de detección y conteo de vacas en la fila de ordeño, utilizando la arquitectura de YOLOv8, logró resultados sobresalientes en términos de precisión y desempeño. Las métricas principales para cada modelo desarrollado y aprobado son las siguientes:

* ***Modelo diurno:*** Precisión del 96.39% en condiciones de luz natural.
* ***Modelo nocturno:*** Precisión del 92.9% en ambientes con poca iluminación.
* ***Modelo híbrido (seleccionado como solución final):***

| **Métrica** | **Valor** |
| --- | --- |
| Recall | 95% |
| mAP50 | 96% |
| mAP@50-95 | 84% |

Estas métricas reflejan un modelo sólido y adaptable para monitorear el flujo de vacas en la fila de ordeño bajo diferentes condiciones de iluminación, cumpliendo con los objetivos establecidos.

***Descripción del Proceso y Costos***

El desarrollo del proyecto siguió la metodología CRISP-DM, asegurando un enfoque estructurado:

1. Preparación y limpieza de datos:

* Se identificaron problemas críticos en el dataset, como imágenes demasiado oscuras o sobreexpuestas, que fueron eliminadas para evitar impactos negativos en el entrenamiento del modelo.
* Se implementaron técnicas de mejora para aumentar la calidad de los datos disponibles.

1. Selección y evaluación de modelos:

* Se investigaron y probaron arquitecturas como YOLOv8 y Faster R-CNN, lo que requirió un tiempo significativo para implementar y evaluar las alternativas.
* Finalmente, se seleccionó YOLOv8 híbrido por su robustez, precisión y capacidad de generalización en condiciones mixtas.

1. Evaluación y ajuste del modelo:

* Se probaron diversos modelos separados para día y noche, antes de optar por el modelo híbrido, ahorrando tiempo y recursos computacionales.

1. Planes de despliegue y mantenimiento:

* Se desarrolló un plan detallado de implementación, que incluye la entrega del repositorio en GitHub con el código completo y un plan de monitoreo y mantenimiento para garantizar el uso continuo y efectivo del sistema.

**La descripción del proceso incluyendo los costos:**

Para llevar a cabo este proyecto, el equipo ha seguido la metodología CRISP-PM. Hemos empezado tratando los datos, preparándonos para los modelos y comprendiendo los, viendo que es lo que podría causar problemas, borrando los outliers (imágenes muy oscuras o muy claras). Luego hemos investigado cuáles son los modelos que podrían ser los más adecuados para nuestro problema, lo cual nos tomó bastante tiempo, causando gran pérdida de tiempo y energía, ya que pasamos bastante tiempo tratando de implementar dos modelos por separado, gastando unidades de cálculo. En este proyecto, los costos se cuentan en tiempo, ya que no hemos pagado para acceder a recursos tipo GPUs. Al haber tratado de entrenar el modelo híbrido y los dos modelos con YOLO v8, y Fast R-CNN, hemos procedido a la evaluación de cada uno de ellos, para llegar a la conclusión que el modelo híbrido de YOLO v8 nano era el más robusto. Para terminar con el proyecto según la metodología, hemos finalmente diseñado un plan de despliegue, que va de la entrega del repositorio GitHub del proyecto, con todo el código, hasta el plan de monitoreo y mantenimiento.

**Desviaciones del Plan Original**A lo largo del proyecto, surgieron varias desviaciones importantes:

* **Identificación individual de vacas**: Inicialmente, se planteó la posibilidad de identificar cada vaca individualmente y rastrear su tiempo de espera. Sin embargo, esta idea se descartó por la falta de un dataset etiquetado a ese nivel de detalle y por los recursos necesarios para entrenar un modelo más avanzado.
* **Separación de modelos por condiciones de iluminación**: Aunque inicialmente se planearon modelos separados para día y noche, finalmente se optó por un modelo híbrido más eficiente y generalizable.
* **Diseño del dashboard**: Gracias al contacto frecuente con el cliente, el diseño del panel fue ajustado continuamente con base en comentarios, logrando una interfaz más funcional y alineada con las necesidades de CAETEC.

**Planes de Implementación**

El modelo final será implementado con los siguientes elementos clave:

* Entrega de un repositorio completo en GitHub con el código documentado para facilitar su uso y mantenimiento.
* Implementación de un dashboard interactivo que permita visualizar métricas clave, como conteo de vacas y posibles aglomeraciones en tiempo real.
* Integración de un plan de monitoreo y mantenimiento para garantizar que el modelo se ajuste a los cambios en el entorno operativo, como nuevas condiciones de iluminación o configuraciones en la fila de ordeño.

**Recomendaciones para Trabajo Futuro**

Para mejorar y extender el alcance del proyecto, se sugieren las siguientes acciones:

* Implementar técnicas avanzadas de data augmentation, como ajustes en el contraste de imágenes, para mejorar la capacidad del modelo en condiciones de iluminación variables.
* Experimentar con arquitecturas más complejas, como versiones avanzadas de YOLOv8 (small o medium), para explorar mejoras adicionales en precisión y capacidad de generalización.
* Obtener acceso a recursos computacionales más robustos, como GPUs de mayor potencia, para entrenar modelos más complejos y realizar ajustes finos con mayor eficiencia.

**Review del proyecto**

***Documentación de la experiencia***

En un proyecto como este donde se tiene el enfoque en la detección de vacas y prevención de aglomeraciones para optimizar la producción lechera, hemos tenido diferentes momentos a lo largo del desarrollo de la solución los cuales han sido trampas potenciales que han afectado a cada fase del proyecto.

Desde el inicio, uno de los principales riesgos que tuvimos durante el entendimiento del negocio fue definir de forma incompleta el problema y los objetivos de negocio que buscamos impactar. Si no se comprende con claridad cómo las aglomeraciones impactan la producción de leche, el proyecto podría centrarse en aspectos secundarios.

En la fase de entendimiento, recolección y preparación de los datos también tuvimos ciertos problemas. Imágenes con mala iluminación llegaron a causar una limitación del desempeño del modelo, por eso se tomó la decisión de eliminar esas imágenes con valores extremos de iluminación ya sean altos o bajos. Por otro lado, también tuvimos cierta problemática en el etiquetado de las imágenes, ya que en una primera iteración de esta tarea, los miembros del equipo sin darnos cuenta etiquetamos las mismas imágenes, por lo que tuvimos que hacerlo de nuevo ya de manera correcta, cada quien con imágenes diferentes.

Durante la etapa de modelado, el seleccionar un modelo inadecuado para los recursos tecnológicos disponibles o para la complejidad de la solución a la que queremos llegar, también puede llevar a resultados desfavorables. Es por ello que probamos con distintos modelos y configuraciones, para evitar la confianza en un solo modelo para no comprometer el desempeño del sistema.

**Enfoque engañoso**

Entrenar dos modelos para un mismo problema fue un error estratégico que incrementó la complejidad del proyecto sin proporcionar beneficios sustanciales. En el futuro, es preferible abordar problemas similares con un único modelo robusto y bien optimizado, utilizando técnicas como el aumento de datos, transferencia de aprendizaje y ajustes hiper paramétricos para abordar las variaciones en los datos (como las condiciones diurnas y nocturnas). Esto reduce la duplicación de esfuerzos y facilita tanto la implementación como el mantenimiento del sistema.

**Lecciones Aprendidas y Recomendaciones para Proyectos Similares**

En el desarrollo de un proyecto de minería de datos como este, es fundamental documentar las mejores prácticas y los aprendizajes obtenidos para facilitar la selección de técnicas adecuadas en condiciones similares en el futuro. Estas recomendaciones, basadas en la experiencia del proyecto, proporcionan una guía clara para enfrentar problemas análogos de manera eficiente y efectiva:

* No tratar de separar lo datos para hacer varios modelos, pero más bien tratar de entender mejor los datos, preparandolos mejor, usando técnicas de data augmentation y modification (por ejemplo cambiar el contraste de las imágenes que sean más oscuras que un cierto límite) para aumentar la varianza y que el modelo pueda más fácilmente entender patrones.
* Pasar más tiempo entrenando otros modelos con otras arquitectura, encontrar el mejor modelo para un cierto problema es intentar otras arquitecturas, y solamente luego mejorarlo.
* En caso de que sea un proyecto importante, en donde la precisión es importante, obtener recursos más grandes, GPUs con mayor potencia, para poder tener tiempo de entrenar modelos más complejos y poder hacer más ajustes.

***Participación de cada miembro del equipo:***

Adrián Matute:

Durante el proyecto, fui elegido como líder del equipo, lo que me permitió tener una visión general de todo el proceso y coordinar las tareas entre los miembros. Tuve una participación constante y activa desde las primeras fases. En el entendimiento del negocio y de los datos, trabajé junto al equipo para definir los objetivos del proyecto, tanto a nivel de negocio como en términos de minería de datos, asegurándonos de que nuestras metas estuvieran alineadas con las necesidades de CAETEC. También colaboré en la elaboración y seguimiento del plan del proyecto, y participé en la primera visita al CAETEC para comprender mejor el contexto y las necesidades de la organización.

En la fase de preparación de datos, mi participación fue clave en varias tareas. A mí y a mis compañeros nos tocó etiquetar aproximadamente 800 fotos, pero debido a un error en la asignación de imágenes, terminé etiquetando el doble de lo planeado. Además, me encargué de limpiar el dataset eliminando imágenes que tenían problemas de luminosidad, ya fueran demasiado claras o demasiado oscuras, lo que podría haber afectado el rendimiento del modelo. También ayudé a separar las imágenes en dos conjuntos: uno para el día y otro para la noche, basado en los nombres de los archivos que indican la hora de captura. Para hacer esto, definimos un umbral de luminosidad que nos permitió clasificar las imágenes de manera más precisa, lo cual resultó ser fundamental para mejorar la calidad del dataset y facilitar el proceso de entrenamiento.

Esta fase fue crucial, ya que aseguramos que los datos con los que trabajamos fueran lo más limpios y organizados posible, lo que nos permitió avanzar con los modelos de manera más efectiva.

En la fase de modelado, me encargué de monitorear y entrenar el modelo de YOLOv8, específicamente el que se utilizaría para las imágenes con buena iluminación, es decir, las tomadas durante el día en el rancho. Además de este, entrené por separado otro modelo, el Faster R-CNN, para evaluar su rendimiento. Sin embargo, este último no logró los resultados esperados en términos de efectividad para abordar la problemática y cumplir con los objetivos que teníamos. Fue un proceso interesante, pero los resultados no fueron los más adecuados para el contexto del proyecto.

A lo largo de estas semanas de trabajo, estuve presente en casi todas las clases y formé parte activa de las reuniones con el socio formador, lo que me permitió mantenerme alineado con los avances del proyecto y con las expectativas del CAETEC.

En cuanto a la evaluación del modelo, decidí realizar una segunda iteración con Faster R-CNN, con el fin de mejorar los resultados. Sin embargo, a pesar de los ajustes realizados, los resultados seguían sin ser los más efectivos en comparación con otros enfoques. Fue una lección importante sobre la necesidad de probar diferentes modelos y ajustar sus parámetros para encontrar la mejor solución.

Finalmente, en la fase de deployment, me encargué de desarrollar el plan de monitoreo y mantenimiento del modelo. Este plan es clave para garantizar que el modelo se mantenga funcional y preciso en el tiempo. También apoyé en la redacción del reporte final, en el que se detallaron los objetivos alcanzados en la fase de "business understanding" y se definieron los próximos pasos para el proyecto, incluyendo posibles mejoras tanto en los modelos como en la metodología de trabajo.

Pablo Martínez:

* + Entrenamiento del modelo diurno y híbrido
* Osvaldo Del Valle:
  + Frontend: dashboard
* Andres Callealta:
  + Entrenamiento de las dos iteraciones de los modelos nocturnos
  + Preparación de los diferentes sets de datos (train y validation)
  + Además de haber puesto en práctica la metodología CRISP-PM a lo largo del proyecto, tuve la oportunidad de escribir partes de todos los proyectos, lo que me permitió entender más en profundidad la metodología, comprendiendo y dándome la cuenta de la importancia que es de seguir un “plano” detallado, sobre todo en proyectos grandes como este, en los cuales hay muchas diferentes fases diferentes, donde el tiempo cuenta mucho, un pequeño retraso en una fase impide a veces poder pasar a la siguiente, tanto al nivel personal que grupal. Me di cuenta de la importancia de hacer las cosas con un orden predefinido, y no lanzarme directamente los ojos cerrados a las partes del proyecto que me gustan más. El proceso aplicado es sobre todo útil en proyectos no personales, donde el cliente interviene especificando sus necesidades o sugerencias. Al nivel personal, la parte que mas me gusto y donde más aprendí, fue entrenar el modelo nocturno, comprendiendo los resultados y metricas para poder
  + Jorge Martínez:
    - Bitácora de seguridad
  + Organizar el repositorio GitHub
  + Entrenamiento de las dos iteraciones del modelo Fast R-CNN
* Más detalle sobre los reportes:

| **Fase** | **ID** | **Producto de trabajo** | **Tipo** | **Involucrados** |
| --- | --- | --- | --- | --- |
| Entendimiento de negocio | 1 | Plan de valor ganado | Documento | Matute y Osvaldo |
| 2 | Plan de mitigación de riesgos | Documento | Pablo |
| 3 | Explicación de metodología utilizada | Documento | Jorge |
| 4 | Criterios de éxito de negocio | Documento | Todos |
| 5 | Criterios de éxito de minería de datos | Documento | Todos |
| 6 | Determinar objetivos de negocio | Documento | Todos |
| 7 | Evaluación de situación | Documento | Todos |
| 8 | Determinar objetivos de minería de datos | Documento | Todos |
| 9 | Plan de proyecto | Documento | Matute |
| Entendimiento de los datos | 10 | Coleccionar datos iniciales | Dataset | Todos |
| 11 | Descripción de datos | Documento | Todos |
| 12 | Exploración de datos | Documento | Todos |
| 13 | Verificación de calidad de datos | Documento | Todos |
| Preparación de datos | 14 | Selección de datos | Desarrollo | Matute |
| 15 | Limpieza de datos | Desarrollo | Jorge |
| 16 | Construcción de datos | Desarrollo | Andres |
| 17 | Integración de datos | Desarrollo | Andres |
| 18 | Formateo de datos | Desarrollo | Matute |
| Modelado | 19 | Seleccionar técnica de modelado | Documento | Osvaldo |
| 20 | Generar diseño de tests | Documento | Matute |
| 21 | Construir modelo | Desarrollo | Todos |
| 21R | Construir modelo | Documento | Jorge |
| 22 | Evaluar modelo | Documento | Andres |
| 26 | Log de consultas a BD | Desarrollo | Osvaldo |
| 30 | Dashboard | Desarrollo | Osvaldo |
| 27 | Script tomar fotos | Desarrollo | Pablo |
| 28 | Script generar predicción | Desarrollo | Pablo |
| 29 | Instalación en raspberry | Desarrollo | Osvaldo y Pablo |
| 33 | Prueba de arquitectura (CAETEC) | Desarrollo | Todos |
| Evaluación | 23 | Evaluar resultados | Documento | Matute |
| 24 | Revisar proceso | Documento | Todos |
| 25 | Determinar pasos siguientes | Documento | Todos |
| Seguridad | 33 | Bitácora de seguridad | Documento | Matute y Jorge |
| Presentación | 31 | Crear presentación | Documento | Todos |
| Finalización de proyecto | 32 | Dinamica de integracion | Dinámica | Todos |
| Deployment | 33 | Despliegue | Documento | Matute y Andres |