



Instituto Tecnológico y de Estudios Superiores de Monterrey

Inteligencia Artificial Avanzada para la Ciencia de Datos

Evaluación

Los concentrados v2

Daniel Queijeiro Albo - A01710441

Diego Alfaro Pinto - A01709971

Diego Isaac Fuentes Juvera - A01705506

Jesus Ramirez Delgado - A01274723

Mauricio Anguiano Juarez - A01703337

Luis Adrián Uribe Cruz - A01783129

Adaptaciones CRISP-DM	3
1.0 Evaluar resultados	3
1.1 Resultados de modelado	3
1.2 Modelos seleccionados	6
2.0 Predicción modelos seleccionados	6
3.0 Análisis del proceso	9
4.0 Trabajo futuro/siguientes pasos	10
4.1 Lista de posibles acciones a futuro	10
A) Mejoras técnicas del modelo	10
B) Expansión de capacidades funcionales	11
C) Evolución hacia un ecosistema de valor real	11
4.2 Decisiones	12

Adaptaciones CRISP-DM

La fase de Evaluación se implementó siguiendo el estándar CRISP-DM, pero con una adaptación específica para el proyecto:

- Se añadió una actividad de validación operativa mediante la ejecución de los modelos seleccionados, con el fin de verificar si las predicciones generadas cumplen los criterios de éxito del negocio.

Esto permitió confirmar, con vacas reales, que las decisiones generadas por el IMR coinciden con el conocimiento experto y apoyan el propósito de mejora genética y manejo del hato.

Además, porque el proyecto se documentó en español. Se realizó la traducción directa de cada actividad de esta fase del inglés al español, manteniendo equivalencia conceptual.

- Evaluate the results ☐ Evaluar resultados.
- Review process ☐ Análisis del proceso.
- Determine next steps ☐ Trabajo futuro/siguientes pasos.

1.0 Evaluar resultados

1.1 Resultados de modelado

Durante la fase de modelado se construyeron modelos para las dimensiones de comportamiento y sanidad que conforman nuestro Índice de Mérito Reproductivo. El proceso fue un ciclo iterativo donde a partir de los resultados basados en los criterios de éxito de las metrics estipuladas para cada modelo, los cuales son:

Modelo	Métrica	Criterio aceptable	Criterio ideal
Comportamiento	Accuracy	≥ 0.80	$\geq 0.85-0.90$
	Precision	≥ 0.75	$\geq 0.80-0.85$
	Recall	≥ 0.70	≥ 0.80
	F1 Score	≥ 0.72	≥ 0.80
Sanidad	% anomalías por fold	$\text{diff} \leq 5$ puntos	$\text{diff} \leq 3$ puntos
	% anomalías global	4%–7%	$\approx 5\%$

	Histograma score	cola visible	outliers muy definidos
--	------------------	--------------	------------------------

Con los resultados logramos tomar decisiones para descartar, optimizar y mejorar o cambiar el enfoque del modelo seleccionado de cada dimensión. Los resultados obtenidos para cada modelo fueron los siguientes:

Modelos de comportamiento

Modelo	Accuracy	Precisión	Recall	F1 Score
Regresión Logística (baseline)	0.7746	0.6639	0.6713	0.6674
Random Forest v1.0	0.909091	0.3333	0.3333	0.3333
Random Forest v2.0	0.8629	0.8910	0.6766	0.7689
Random Forest v2.1 (Optimizado)	0.8699	0.8964	0.6943	0.7824
MLP Optimizado	0.8505	0.8068	0.7351	0.7682

El experimento comparó cuatro modelos para identificar vacas inquietas. La **Regresión Logística** sirvió como punto de partida, pero quedó por debajo de los criterios mínimos en todas las métricas, funcionando únicamente como referencia comparativa.

La primera versión de **Random Forest (v1.0)** mostró resultados artificialmente altos en accuracy pero con métricas inestables y poco confiables debido a un dataset pequeño y desbalanceado, por lo que fue descartada y el modelo se reconstruyó con más datos.

La versión **Random Forest v2.0** logró un desempeño sólido con alta precisión, pero aún dejó escapar casos reales (recall bajo), mostrando oportunidad de mejora.

La versión **Random Forest v2.1**, optimizada mediante búsqueda de hiper parámetros, presentó el mejor equilibrio global entre precisión, recall y estabilidad entre folds, convirtiéndose en el modelo más maduro y recomendable del enfoque basado en árboles.

Finalmente, el **MLP (red neuronal)** prioriza la detección de casos inquietos y fue el único que superó el criterio mínimo de recall, siendo útil cuando la prioridad es capturar el mayor número posible de positivos aunque implique sacrificar algo de precisión.

Modelos de sanidad

Modelo	Variación entre folds	% Anomalías global	Histograma (distribución)	PCA (separación espacial)
Modelo base (DBSCAN v1.0)	1.98 p.p. (mejor estabilidad)	4.97% (óptimo)	Cola clara + corte 5%	Clúster compacto + outliers limpios
Isolation Forest 1.0	60 p.p. (muy inestable – no cumple)	16.11% (fuera del rango – no cumple)	Cola marcada pero con extremos amplios (0–60%) — separación inconsistente	Outliers presentes pero dispersos; separación aceptable pero no homogénea
Isolation Forest 2.0	~4 p.p. (inestable vs otros)	7.22% (sobre-detección)	Frontera difusa - menos separación	Separación visual aceptable
Isolation Forest 2.1	2.42 p.p. (ideal)	5.03% (equilibrado)	Corte limpio top 5%, más definido que v2.0	Cluster normal más compacto + outliers aislados

El experimento evaluó 4 enfoques no supervisados para detectar anomalías sanitarias: DBSCAN, Isolation Forest 1.0, Isolation Forest 2.0 e Isolation Forest 2.1. El análisis buscó medir estabilidad, claridad en la separación de casos atípicos e interpretabilidad de resultados.

DBSCAN v1.0 sirvió como referencia sólida. Mostró la mayor estabilidad entre pliegues, detectó anomalías en el rango ideal (~5 %) y evidenció una segmentación limpia de outliers, funcionando como curva base para comparar el desempeño del resto de modelos.

Isolation Forest v1.0, entrenado con un dataset pequeño, resultó inestable y sobreajustado: la tasa de anomalías fluctuó drásticamente entre folds y el promedio global (16.11 %) se alejó del objetivo. Aunque sugería cierta separación visual, sus resultados eran inconsistentes, por lo que se descartó y se reentrenó con una base ampliada.

Isolation Forest v2.0 ofreció mejoras, pero mostró sobre-detección (7.22 % de anomalías) y menor definición geométrica, lo que confirmó la necesidad de refinar estabilidad y calibración del umbral.

Finalmente, **Isolation Forest v2.1** alcanzó el equilibrio deseado: proporción ideal de anomalías (~5 %), variabilidad controlada entre folds (~2.42 p.p.), histograma más limpio y mayor separación espacial de atípicos en PCA. Esto lo convierte en el modelo más maduro e interpretable, recomendado para despliegue operativo y como base para futuros métodos avanzados.

El objetivo de esta fase es contrastar el desempeño de los modelos contra los **criterios de éxito definidos al inicio del proyecto**, con el fin de justificar la selección de las versiones **v2.1** como modelos finales para su despliegue.

1.2 Modelos seleccionados

Durante iteraciones previas se exploraron modelos base, versiones iniciales y optimizaciones progresivas. Sin embargo, no todos alcanzaron los umbrales requeridos para considerarse adecuados para uso operativo, ya sea por sensibilidad insuficiente, inestabilidad entre ciclos o sobre-detección de casos.

Las versiones **v2.1** tanto para comportamiento como para sanidad representan las primeras versiones que:

1. Cumplen los criterios aceptables o ideales en **al menos tres métricas clave**,
2. muestran **estabilidad** ante validaciones cruzadas,
3. mantienen **comportamiento consistente** visual y estadísticamente,
4. superan los problemas observados en versiones anteriores (dataset insuficiente, sobre detección, baja sensibilidad).

Por lo anterior, ambas v2.1 son seleccionadas como modelos candidatos finales para integración y despliegue.

2.0 Predicción modelos seleccionados

Una vez con los modelos seleccionados, se seleccionaron 3 vacas no incluidas en el dataset de entrenamiento que cumplen con características de las 3 dimensiones (comportamiento, sanidad, producción) que clasificamos, como excelente, medianamente buena, mala. Los resultados esperados son:

Id Vaca	IMR Esperado
1554	Retener/Reproducir
1634	Supervisar
8755	Descartar/Secado

Para el Índice de Mérito Reproductivo es importante recordar cómo está construido, el cual consiste de la siguiente forma:

Primero se **descuenta el efecto del ambiente** en cada sesión de ordeño. Para ello se calcula una producción esperada en función del contexto (robot / hora / mes) y se define la producción ajustada como:

$$ProduccionAjustada_i = ProduccionObservada_i - \bar{x}(Produccion | robot \times hora \times mes)$$

Luego, el **Mérito Productivo de cada vaca** se obtiene agregando sus producciones ajustadas a lo largo del periodo:

$$MeritoProductivo_i = \frac{1}{N_i} \sum_{j=1}^{N_i} ProduccionAjustada_{i,j}$$

Donde:

i = vaca

j = sesión de ordeño de esa vaca

N_i = número de ordeños que tiene esa vaca en el período

Finalmente, este mérito se integra en el **Índice de Mérito para la Retención (IMR)**, combinándolo con los riesgos de comportamiento y sanidad mediante una versión estandarizada:

$$IMR_i = w_G \cdot Z(MeritoProductivo_i) - w_C \cdot Z(RiesgoComportamiento_i) - w_S \cdot Z(RiesgoSanidad_i)$$

con pesos:

- w_G = peso del mérito productivo (peso seleccionado: 0.5)
- w_C = peso del riesgo de comportamiento (peso seleccionado: 0.2)
- w_S = peso del riesgo sanitario (peso seleccionado 0.3)

Es importante mencionar que los pesos seleccionados son en base a lo que tenemos más certeza o datos ciertos en nuestros modelos, dando mayor influencia a el modelo de comportamiento dados los datos que nos confirman con mayor precisión el comportamiento de una vaca. Estos pesos se pueden ajustar si se quiere dar más influencia o menos a alguna de las dimensiones.

El rango de toma de decisión es el siguiente:

Rango del IMR (basado en percentiles poblacionales)	Condición estadística	Decisión recomendada	Interpretación operativa
IMR \geq p75	Top 25% del hato	Retener / Reproducir	Alta prioridad genética y productiva - candidata para inseminación o mantenimiento prolongado.
p40 \leq IMR < p75	Segmento medio (p40–p75)	Supervisar / Manejo dirigido	Vacas de comportamiento intermedio - conservar, monitorear y optimizar manejo antes de decidir reproducción o descarte.
IMR < p40	Bottom 40% del hato	Descartar / Secado / Venta	Bajo desempeño relativo - adecuada para retiro, secado o priorizar reemplazo.

Con esto establecido, los resultados del IMR fueron los siguientes:

Id Vaca	IMR Esperado	IMR Obtenido	Resultado obtenido
1554	Retener/Reproducir	Retener/Reproducir	IMR: 3.2629639425555865
1634	Supervisar	Supervisar	IMR: 0.284701617131891
8755	Descartar/Secado	Descartar/Secado	IMR: -3.3294829705559494

Los resultados de integración operacional confirmaron el desempeño esperado del índice calculado. Las tres vacas evaluadas mostraron coherencia total entre la decisión esperada y la decisión generada automáticamente por el IMR. La vaca 1554 fue correctamente clasificada como candidata a Retener/Reproducir con un IMR elevado (3.26), la vaca 1634 fue asignada a la categoría de Supervisión/Manejo dirigido con un IMR intermedio (0.28), y la vaca 8755 fue correctamente identificada como caso de Descarte/Secado, reflejado por un IMR negativo (-3.33). Esta alineación entre los valores esperados y las decisiones obtenidas valida el funcionamiento del modelo en condiciones reales, demostrando capacidad de priorización adecuada y correspondencia con criterios productivos del sistema.

El código de integración se encuentra disponible en el [repositorio del proyecto](#). En la raíz del repositorio el archivo integration_v2.py

3.0 Análisis del proceso

Durante el desarrollo del proyecto se siguió la metodología CRISP-DM, aplicando sus fases principales y realizando diversas adaptaciones para ajustarse a las necesidades del equipo y del propio proyecto. Entre estas adaptaciones se incorporó un plan de valor ganado, con el propósito de mantener un control adecuado del progreso y una distribución clara de responsabilidades. Cada actividad fue desglosada en tareas más pequeñas y asignada a distintos integrantes del equipo para asegurar una ejecución organizada.

Asimismo, en la fase de *Business Understanding* se modificó la actividad “Costs and Benefits”, debido a que el proyecto no persigue un objetivo monetario, sino un propósito académico orientado al beneficio del Tecnológico de Monterrey.

En la fase de *Data Understanding* se añadió un componente adicional en el que se documentó el tratamiento y la anonimización de los datos, cumpliendo con los requisitos del socio formador y respondiendo al contexto del proyecto.

Finalmente, durante la fase de *Modeling*, el equipo acordó que al menos un modelo de cada propuesta sería implementado. Cada iteración, basada en técnicas de machine learning distintas, fue desarrollada por un miembro diferente del equipo, promoviendo así la distribución equitativa del trabajo y la diversidad en los enfoques de modelado.

El desarrollo del proyecto permitió transitar de exploración inicial, construcción de variables, experimentación con modelos y validación operativa hacia un sistema funcional capaz de generar indicadores para la toma de decisiones en el hato. El proceso resultó altamente iterativo y enriquecedor: cada fase aportó aprendizaje técnico y conocimiento contextual sobre el comportamiento animal y el funcionamiento del sistema robotizado.

Los resultados obtenidos satisfacen los criterios de éxito establecidos tanto para la minería de datos como para los objetivos de negocio. Desde una perspectiva técnica, los modelos alcanzaron niveles de desempeño alineados con los umbrales definidos y demostraron estabilidad, interpretabilidad y capacidad predictiva. Desde la perspectiva de negocio, el sistema logró identificar vacas inquietas, estimar riesgo sanitario y calcular un índice combinado de mérito reproductivo, permitiendo clasificar animales en categorías accionables (retener, supervisar, descartar). Esto se traduce directamente en indicadores útiles para la mejora genética y el manejo operativo del hato, atendiendo plenamente el objetivo estratégico del proyecto.

El resultado fue revisado y presentado al stakeholder académico principal, Dr. Ivo Neftalí Ayala García, quien expresó satisfacción con la calidad técnica del modelo, la claridad de su interpretación y la relevancia de los hallazgos para la industria lechera. Su retroalimentación confirma que los resultados cumplen los criterios de éxito del negocio, especialmente la identificación de:

- Conductas asociadas con mayor producción,
- Patrones fisiológicos vinculados con riesgo sanitario, y

- Rasgos compatibles con el sistema robotizado.

Finalmente, el equipo concluye que el proyecto ha sido exitoso en términos técnicos, funcionales y de aceptación por parte del usuario experto, y se ha acordado establecer un plan de monitoreo y mejora continua. Este contempla recalibración periódica de percentiles globales del IMR, actualización del mérito productivo conforme se incorporan más animales y evaluación continua del rendimiento de los modelos, asegurando sostenibilidad, trazabilidad y evolución del sistema a largo plazo.

4.0 Trabajo futuro/siguientes pasos

4.1 Lista de posibles acciones a futuro

Debido a que el proyecto concluye el 5 de diciembre, se proponen las siguientes líneas de continuidad y expansión para fortalecer el sistema desarrollado para el CAETEC:

A) Mejoras técnicas del modelo

1. **Reentrenamiento periódico con datos históricos.**

Conforme el rancho continúe acumulando registros de ordeño y nuevas cohortes de vacas, se recomienda implementar ciclos trimestrales o semestrales de actualización de los modelos. Esto permitiría capturar variaciones estacionales, cambios de manejo y efectos fisiológicos del hato a lo largo del tiempo.

2. **Validación inter-robot y transferencia.**

El proyecto se desarrolló exclusivamente utilizando datos del robot BMC-2. Futuros trabajos pueden evaluar el desempeño de los modelos en los robots BMC-1 y BMC-3, o incluso en establos externos. Esta validación cruzada permitirá determinar si los algoritmos generalizan de forma robusta a distintas condiciones de operación, infraestructura y manejo animal.

3. **Optimización automatizada de hiper parámetros.**

Si bien en este proyecto se realizaron ajustes manuales basados en experimentación iterativa, futuras investigaciones podrían incorporar Bayesian Optimization o AutoML para acelerar el descubrimiento de arquitecturas óptimas.

4. **Enriquecimiento multivariable del modelado.**

Futuras iteraciones pueden integrar nuevos factores predictivos tales como:

- Número de partos,
- Edad de la vaca,
- Lactancias previas.
- Historial sanitario.
- Estacionalidad climática,
- Composición del alimento.

- Estrés térmico medido por THI (Temperature Humidity Index). La incorporación de estas dimensiones permitiría modelar comportamiento, producción y salud con un contexto biológico y productivo más completo.

B) Expansión de capacidades funcionales

1. Integración de visión computacional.

El rancho cuenta con más de 50,000 imágenes captadas por cámaras del sistema robotizado. Un módulo de análisis visual (CNN, EfficientNet, YOLO) permitiría detectar lesiones, inflamación en cuartos, malformaciones o comportamientos anormales documentados visualmente, complementando la predicción basada en indicadores numéricos.

2. Sistema multicapa de alerta y priorización inteligente.

En lugar de alertas binarias, se puede implementar un sistema de niveles (bajo, medio, alto) que incorpore severidad clínica, frecuencia histórica, vulnerabilidad genética y progresión temporal, permitiendo decisiones de manejo más estratégicas.

3. Predicción individual de producción lechera futura.

Modelos de series temporales como ARIMA, SARIMA, Prophet o redes LSTM pueden anticipar caídas de producción o ciclos de recuperación, ayudando a priorizar dieta, manejo o intervenciones sanitarias antes de que el problema sea visible.

C) Evolución hacia un ecosistema de valor real

1. Despliegue más allá del ejecutable actual.

El plan actual contempla un ejecutable standalone. Una evolución natural es desarrollar:

- una aplicación web con dashboard dinámico,
- una API que permita automatizar decisiones en tiempo real,
- integración con IoT dentro del establo,
- o una aplicación móvil para veterinarios en campo.

Estas versiones permitirían reportes automáticos diarios, comparativos por lote, seguimiento clínico y recomendaciones personalizadas.

2. Sistema de aprendizaje continuo (MLOps rural).

Implementar pipelines de retraining, validación, versionado de modelos, auditoría de decisiones y monitoreo de drift. Esto convertiría la solución en una plataforma viva y escalable.

3. Expansión científica y académica.

Dado que los resultados entusiasmaron al stakeholder principal, el Dr. Ivo Ayala, una línea

estratégica futura es formalizar el proyecto como:

- Caso de estudio institucional.
- Publicación científica.
- Spin-off tecnológico.
- Proyecto PhD/MSc de continuidad en agro-IA aplicada.

4.2 Decisiones

Dados los resultados positivos de la fase de modelado, los modelos seleccionados con respaldo de los resultados y las predicciones integradas al mérito de producción que dan como resultado el Índice de Mérito Reproductivo, el equipo decidió priorizar la entrega de un sistema funcional, verificable y listo para implementación inmediata, alineado con los criterios de éxito del cliente y el calendario del proyecto.

En consecuencia, antes del 5 de diciembre se acordó entregar:

- Modelos de comportamiento y sanidad completamente entrenados, calibrados y serializados, permitiendo inferencia directa sin necesidad de reentrenamiento.
- Un ejecutable (.exe) que encapsula el proceso de integración y cálculo del Índice de Mérito Reproductivo (IMR), conforme al plan de despliegue establecido. Este componente permite que el rancho utilice el modelo sin requerir un entorno de desarrollo o conocimientos técnicos avanzados.
- Código fuente y un manual de usuario, asegurando trazabilidad, reproducibilidad y continuidad del proyecto.

Adicionalmente, se definió un plan de monitoreo y mejora continua, el cual contempla horas de apoyo técnico para ajustes solicitados a la aplicación que no superen dos horas de trabajo.

Las decisiones tomadas se enfocan en entregar valor tangible y facilitar adopción operativa.