



*Master in Management + Analytics*

# Optimización Data-Driven del Número de Transferencias Embrionarias en Clonación Equina

**Autoría:** Picighelli Giacomelli, Augusto

**Año:** 2025

## ¿Cómo citar este trabajo?

Picighelli Giacomelli, A. (2025) *"Implementar un modelo de predicción del máximo segmento a alcanzar por un usuario en aplicación de q-commerce"*. [Tesis de maestría. Universidad Torcuato Di Tella]. Repositorio Digital Universidad Torcuato Di Tella.

<https://repositorio.utdt.edu/handle/20.500.13098/13749>

El presente documento se encuentra alojado en el **Repositorio Digital de la Universidad Torcuato Di Tella** bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 4.0 Internacional

**Dirección:** <https://repositorio.utdt.edu>



**UNIVERSIDAD  
TORCUATO DI TELLA**

MASTER IN MANAGEMENT + ANALYTICS

**Optimización Data-Driven del Número de  
Transferencias Embrionarias en Clonación  
Equina**

**TESIS**

Augusto Picighelli Giacomelli

Mayo 2025

Tutor: Jorge Néstor Domínguez

## Resumen

La clonación equina mediante transferencia nuclear somática (SCNT) ha permitido la multiplicación genética de ejemplares de alto valor deportivo y comercial. Dentro de este proceso, la asignación de embriones a yeguas receptoras constituye un punto crítico de eficiencia técnica y económica. Tradicionalmente, se asigna un número fijo de embriones por clon solicitado, buscando maximizar la probabilidad de éxito en un contexto de alta incertidumbre biológica. Sin embargo, esta estrategia estándar no ha sido previamente evaluada desde una perspectiva analítica rigurosa que combine datos históricos, modelización predictiva y simulación de escenarios.

El objetivo de esta tesis fue analizar críticamente el protocolo actual y determinar si existe un número de transferencias más eficiente. Para ello, se trabajó sobre una base de datos histórica de embriones transferidos en un centro de clonación equina de referencia. Se aplicaron técnicas de aprendizaje automático (XGBoost) para construir un modelo predictivo de la probabilidad de éxito por embrión, considerando variables como el tipo de embrión, la calidad, el día de evolución y el estado de conservación (fresco o vitrificado). A partir de las probabilidades estimadas, se desarrollaron simulaciones Montecarlo que permitieron evaluar diferentes escenarios operativos, considerando las restricciones reales del sistema como la cantidad limitada de yeguas receptoras.

Los resultados fueron contundentes: reducir el número máximo de transferencias por clon un 29% maximiza la rentabilidad neta, disminuye el costo por clon entregado y mejora el índice de beneficio comercial (CBR). Esta política de asignación optimizada también contribuye a reducir el número de excedentes de clones no deseados, mejora la eficiencia percibida por los clientes y diversifica el riesgo reproductivo entre líneas genéticas. El estudio demuestra que, aunque no es posible manipular las características biológicas individuales de los embriones, sí se puede optimizar su asignación de manera estratégica.

Adicionalmente, el análisis reveló diferencias en la probabilidad de éxito según características del embrión, validando la importancia de incorporar herramientas analíticas más precisas en la toma de decisiones operativas.

La evidencia presentada muestra que la gestión reproductiva en clonación equina puede beneficiarse significativamente de un enfoque basado en datos, simulaciones estocásticas y modelado predictivo. Estos resultados no solo mejoran el desempeño técnico y económico del sistema actual, sino que también ofrecen una base metodológica sólida para optimizar decisiones en otros procesos biotecnológicos sujetos a alta incertidumbre.

## **Abstract**

Equine cloning through somatic cell nuclear transfer (SCNT) enables the genetic multiplication of high-value horses. A critical point in this process is the allocation of embryos to recipient mares, traditionally assigning a fixed number per requested clone to maximize success under high biological uncertainty. This thesis critically evaluates whether the current standard number of embryo transfers optimizes outcomes. Using historical data from a leading cloning center, a predictive model based on machine learning (XGBoost) was developed to estimate embryo success probabilities, incorporating variables such as embryo type, quality, development day, and conservation status. Montecarlo simulations were applied to assess operational scenarios considering real system constraints, particularly the limited number of recipients.

The results show that reducing the maximum number of transfers per clone by 29% maximizes net profitability, lowers the cost per delivered clone, and improves the commercial benefit ratio (CBR). Optimizing embryo allocation also decreases surplus clones, enhances perceived client efficiency, and diversifies reproductive risk across genetic lines. Additionally, differences in success probabilities according to embryo characteristics reinforce the value of data-driven decision-making.

This study demonstrates that reproductive management in equine cloning can significantly benefit from predictive modeling and stochastic simulation, offering a robust methodological basis to optimize decisions in other biotechnological processes under uncertainty.

# Contenido

|                                                                                                                     |    |
|---------------------------------------------------------------------------------------------------------------------|----|
| Resumen.....                                                                                                        | 2  |
| Abstract .....                                                                                                      | 3  |
| Contenido .....                                                                                                     | 4  |
| Índice de Tablas.....                                                                                               | 7  |
| Índice de Figuras .....                                                                                             | 8  |
| 1. Introducción .....                                                                                               | 9  |
| 1.1. Contexto.....                                                                                                  | 9  |
| 1.1.1. Importancia de la clonación equina en el polo y la genética deportiva .....                                  | 9  |
| 1.1.2. Casos emblemáticos en la clonación equina argentina .....                                                    | 10 |
| 1.1.3. El proceso detallado de clonación equina .....                                                               | 10 |
| 1.1.4. Disponibilidad limitada de yeguas receptoras: Factor limitante clave en el proceso de clonación equina ..... | 11 |
| 1.1.5. Requisitos fisiológicos y reproductivos de las yeguas receptoras .....                                       | 11 |
| 1.1.6. Limitaciones cuantitativas y temporales.....                                                                 | 12 |
| 1.1.7. Complejidades operativas adicionales y manejo de lotes embrionarios.....                                     | 12 |
| 1.1.8. Impacto comercial y económico de la escasez de receptoras .....                                              | 12 |
| 1.1.9. Aspectos económicos y rentabilidad del proceso.....                                                          | 13 |
| 1.1.10. Necesidad estratégica de optimización operativa .....                                                       | 14 |
| 1.2. Problema.....                                                                                                  | 14 |
| 1.2.1. Limitaciones del modelo actual de transferencia embrionaria .....                                            | 14 |
| 1.2.2. Uso fijo del número de embriones transferidos por clon.....                                                  | 15 |
| 1.3. Objetivo .....                                                                                                 | 17 |
| 1.3.1. Pregunta de investigación.....                                                                               | 17 |
| 1.3.2. Objetivo general.....                                                                                        | 17 |
| 1.3.3. Objetivos específicos.....                                                                                   | 17 |
| 1.3.4. Justificación y resultados esperados.....                                                                    | 18 |
| 2. Marco teórico.....                                                                                               | 19 |
| 2.1. Fundamentos de la clonación equina y transferencia embrionaria.....                                            | 19 |
| 2.2. Eficiencia reproductiva y calidad embrionaria .....                                                            | 20 |
| 2.3. Optimización basada en datos: modelos analíticos y toma de decisiones .....                                    | 21 |
| 2.3.1. Técnicas de aprendizaje automático aplicadas a la predicción de eficiencia reproductiva .....                | 22 |
| 2.3.2. Simulación Montecarlo en contextos biotecnológicos: aplicaciones a la clonación equina .....                 | 23 |

|        |                                                                                                  |    |
|--------|--------------------------------------------------------------------------------------------------|----|
| 2.4.   | Integración conceptual: hacia un modelo predictivo y operativo de transferencia embrionaria..... | 24 |
| 3.     | Datos .....                                                                                      | 24 |
| 3.1.   | Origen de la información y proceso de unificación .....                                          | 25 |
| 3.2.   | Composición del dataset y definición de variables.....                                           | 26 |
| 3.3.   | Control de datos faltantes e inconsistencias .....                                               | 27 |
| 3.4.   | Exploración descriptiva .....                                                                    | 27 |
| 3.4.1. | Distribución de transferencias por temporada:.....                                               | 28 |
| 3.4.2. | Calidad de los embriones transferidos: .....                                                     | 29 |
| 3.4.3. | Distribución de embriones según estadio (desarrollo):.....                                       | 30 |
| 3.4.4. | Estado de conservación al momento de la transferencia:.....                                      | 30 |
| 3.4.5. | Día de evolución del embrión previo a la transferencia: .....                                    | 31 |
| 3.4.6. | Distribución del promedio de transferencias por clon entregado.....                              | 32 |
| 3.5.   | Cobertura temporal y validez .....                                                               | 33 |
| 4.     | Metodología .....                                                                                | 33 |
| 4.1.   | Modelo predictivo .....                                                                          | 33 |
| 4.1.1. | Variable Objetivo (TRANSFER_STATUS): .....                                                       | 33 |
| 4.1.2. | Selección de Variables Predictoras .....                                                         | 34 |
| 4.1.3. | Limpieza y Tratamiento de Datos .....                                                            | 34 |
| 4.1.4. | Balanceo de Clases .....                                                                         | 34 |
| 4.1.5. | Modelado y Experimentación .....                                                                 | 34 |
| 4.2.   | Diversificación de líneas genéticas .....                                                        | 35 |
| 4.2.1. | Planteo del problema y enfoque económico .....                                                   | 36 |
| 4.2.2. | Implementación con Simulación Montecarlo .....                                                   | 37 |
| 5.     | Resultados .....                                                                                 | 38 |
| 5.1.   | Resultados del modelo predictivo .....                                                           | 38 |
| 5.1.1. | Desempeño general de los modelos.....                                                            | 39 |
| 5.1.2. | Combinaciones Óptimas para Maximizar la Probabilidad de Nacimiento.....                          | 45 |
| 5.1.3. | Implicancias para la gestión estratégica .....                                                   | 46 |
| 5.2.   | Simulación de escenarios de reducción del número de transferencias .....                         | 46 |
| 5.2.1. | Lógica de la simulación .....                                                                    | 46 |
| 5.2.2. | Clones intentados y exitosos.....                                                                | 47 |
| 5.2.3. | Costo por clon exitoso.....                                                                      | 47 |
| 5.2.4. | Rentabilidad neta y retorno global .....                                                         | 48 |
| 5.2.5. | Eficiencia integral: Commercial Benefit Ratio (CBR) .....                                        | 49 |
| 5.2.6. | Implicancias prácticas .....                                                                     | 50 |

|        |                                                                                |    |
|--------|--------------------------------------------------------------------------------|----|
| 5.3.   | Evaluación económica y estratégica .....                                       | 51 |
| 5.3.1. | Síntesis de hallazgos.....                                                     | 51 |
| 5.3.2. | Restricciones reales y criterios de decisión.....                              | 51 |
| 5.3.3. | Implicancias estratégicas .....                                                | 51 |
| 6.     | Discusión .....                                                                | 52 |
| 6.1.   | Redefinición de la lógica operativa en contextos de alta incertidumbre .....   | 52 |
| 6.2.   | Aportes del enfoque analítico al negocio de la clonación equina .....          | 52 |
| 6.3.   | Limitaciones del estudio .....                                                 | 52 |
| 6.4.   | Lineamientos para futuras investigaciones .....                                | 53 |
| 6.5.   | Reflexión final .....                                                          | 53 |
| 7.     | Conclusión .....                                                               | 54 |
| 7.1.   | Síntesis de objetivos y hallazgos .....                                        | 54 |
| 7.2.   | Contribuciones prácticas y metodológicas .....                                 | 54 |
| 7.3.   | Implicancias estratégicas y de gestión.....                                    | 55 |
| 7.4.   | Límites del trabajo y perspectivas futuras.....                                | 55 |
| 7.5.   | Reflexión final .....                                                          | 55 |
|        | Referencias.....                                                               | 57 |
|        | Apéndice A. El proceso de clonación equina: Pasos para clonar un caballo ..... | 59 |

Índice de Tablas

Tabla 1. Ilustración resumida del dataset. .... 28

Tabla 2. Matriz de confusión del modelo de Árbol de Decisión. .... 39

Tabla 3. Matriz de confusión del modelo Random Forest sin optimización ..... 40

Tabla 4. Matriz de confusión del modelo Random Forest optimizado ..... 41

Tabla 5. Matriz de confusión del modelo XGBoost sin optimizar ..... 42

Tabla 6. Matriz de confusión del modelo XGBoost optimizado ..... 44

Tabla 7. Combinaciones Óptimas para Maximizar la Probabilidad de Nacimiento ..... 45

## Índice de Figuras

|                                                                                                                        |    |
|------------------------------------------------------------------------------------------------------------------------|----|
| Ilustración 1 Cantidad de embriones transferida por temporada (2021-2025). .....                                       | 28 |
| Ilustración 2. Distribución de la calidad de los embriones transferidos. ....                                          | 29 |
| Ilustración 3. Distribución de embriones transferidos según su estadio de desarrollo.....                              | 30 |
| Ilustración 4. Estado de conservación de los embriones transferidos (frescos vs. vitrificados). 31                     |    |
| Ilustración 5. Distribución de transferencias según la edad del embrión al momento de la implantación.....             | 31 |
| Ilustración 6. Histograma de promedio de transferencias por clon entregado (temporadas 21/22 a 23/24). ....            | 32 |
| Ilustración 7. Curva ROC del modelo de Árbol de Decisión. ....                                                         | 40 |
| Ilustración 8. Curva ROC del modelo Random Forest sin optimización. ....                                               | 41 |
| Ilustración 9. Curva ROC del modelo Random Forest optimizado. ....                                                     | 42 |
| Ilustración 10. Curva ROC del modelo XGBoost sin optimizar. ....                                                       | 43 |
| Ilustración 11. Curva ROC del modelo XGBoost optimizado.....                                                           | 44 |
| Ilustración 12. Relación entre clones intentados y clones exitosos por escenario (1 a 10 transferencias). ....         | 47 |
| Ilustración 13. Índice de Costo por Clon Exitoso vs. Transferencias por Clon Contratado .....                          | 48 |
| Ilustración 14. Curva de ingresos, costos y rentabilidad neta relativa según cantidad de transferencias por clon ..... | 49 |
| Ilustración 15. Índice Relativo del Commercial Benefit Ratio vs. Transferencias Máximas por Clon .....                 | 50 |

# 1. Introducción

La industria equina vinculada al polo y la mejora genética ha experimentado en las últimas dos décadas un avance significativo gracias a la adopción de distintas biotecnologías reproductivas, siendo la clonación equina y la ICSI (inyección intracitoplasmática de espermatozoides) dos de las más destacadas. Si bien ambas han demostrado su eficacia para incrementar la disponibilidad de caballos de alto rendimiento, continúan enfrentando limitaciones prácticas que frenan su potencial. En particular, la escasez de yeguas que puedan prestar su vientre para la escalabilidad de estas técnicas constituye un obstáculo crítico para el crecimiento de estos programas, ya que determina la capacidad efectiva de producir más ejemplares y, por ende, de atender la creciente demanda de estos servicios.

## 1.1. Contexto

### 1.1.1. Importancia de la clonación equina en el polo y la genética deportiva

El polo de alto rendimiento, disciplina de élite con profundas raíces en la cultura ecuestre argentina, exige caballos que combinen una extraordinaria capacidad atlética, estabilidad temperamental y elevada resistencia física. La biotecnología reproductiva ha emergido como una herramienta clave para satisfacer estos requerimientos, destacándose la clonación equina como un avance revolucionario que permite replicar individuos con fenotipos deportivos sobresalientes. Esta tecnología ha viabilizado la preservación y difusión de genéticas excepcionales, contribuyendo al perfeccionamiento sistemático de líneas competitivas (Galli et al., 2013; Suvá et al., 2023).

La clonación mediante transferencia nuclear somática ha transformado los paradigmas tradicionales de selección genética, posibilitando la reproducción exacta de individuos ya probados, lo que minimiza la incertidumbre inherente a los métodos convencionales y acelera la consolidación de poblaciones con altos estándares de desempeño (Galli et al., 2013). Esta ventaja es especialmente significativa en el polo, donde el diferencial competitivo puede depender críticamente de la calidad genética del equino.

En combinación con otras biotecnologías reproductivas como la transferencia de embriones, la inseminación artificial y la inyección intracitoplasmática de espermatozoides (ICSI), la clonación ha potenciado el impacto global del mejoramiento genético equino. Estas técnicas permiten disociar la reproducción de las restricciones fisiológicas de las yeguas atletas, maximizando su vida útil en la competencia sin sacrificar su valor reproductivo (Vélez & Hinrichs, 2011; Galli et al., 2013).

Particularmente, la ICSI se ha consolidado como la técnica más eficaz para la fertilización in vitro en equinos, dada la limitada eficiencia de la FIV convencional en esta especie (Grady & Hinrichs, 2016; Oliveira et al., 2024). Su aplicación ha permitido generar embriones viables a partir de semen de escasa calidad o incluso de ejemplares fallecidos, ampliando las posibilidades de manejo genético y conservación de líneas valiosas (Oliveira et al., 2024).

Por otro lado, las tasas de éxito de las tecnologías in vitro aún presentan variabilidad en comparación con los métodos tradicionales. La eficiencia del desarrollo embrionario y la estructura del blastocisto in vitro pueden diferir significativamente de los generados in vivo, evidenciando una menor compactación del ICM y una dispersión de las células del epiblasto, lo cual podría explicar tasas más altas de pérdida embrionaria temprana (Umair et al., 2023).

Desde una perspectiva operativa, estas tecnologías han demostrado ser especialmente útiles en la reproducción de yeguas en actividad deportiva. A través de la recolección de ovocitos mediante punción transvaginal y su fertilización posterior por ICSI, es posible mantener la genética en circulación sin comprometer la performance en cancha, lo que resulta en una estrategia reproductiva de alto retorno tanto económico como deportivo (Carrillo Sánchez, 2021; Galli et al., 2013).

Sin embargo, la implementación extensiva de estas técnicas plantea desafíos bioéticos y regulatorios. El bienestar animal, la trazabilidad genética y la equidad en la competencia son aspectos que requieren un abordaje riguroso y transparente para asegurar la aceptación social de estas prácticas y su sostenibilidad a largo plazo (Galli, 2013).

### **1.1.2. Casos emblemáticos en la clonación equina argentina**

Dos casos particularmente emblemáticos en la historia reciente del polo ilustran claramente el potencial transformador de la clonación equina. El primero corresponde a la yegua Dolfina Cuartetera, ampliamente reconocida por su desempeño excepcional en competencias internacionales de alto nivel. Cuartetera, propiedad del destacado jugador Adolfo Cambiaso, logró múltiples reconocimientos por su capacidad de juego única y desempeño constante en la cancha. La decisión de clonarla permitió preservar su genética incomparable, garantizando la continuidad de su excepcional linaje a través de múltiples generaciones clonadas. Actualmente, sus clones compiten regularmente en las principales competencias internacionales, manteniendo un rendimiento consistente y confirmando así la eficacia de la técnica de clonación en la preservación y potenciación genética del polo de élite (The Washington Post, 2023).

Otro caso significativo es el de la yegua Silvina Luna, que representa un hito clave al ser el primer clon equino desarrollado específicamente para el polo en Argentina. Nacida en 2013 gracias al trabajo del laboratorio Kheiron Biotech, Silvina Luna estableció un precedente tecnológico esencial, validando científicamente el uso práctico y efectivo de la clonación en la producción equina deportiva en Sudamérica. Su nacimiento no solo destacó por su relevancia científica, sino también por sus implicaciones comerciales, permitiendo al laboratorio consolidar su liderazgo tecnológico regional y expandir sus operaciones comerciales, ofreciendo servicios de clonación a nivel internacional. Este logro tuvo un impacto significativo en la percepción pública de la clonación, generando un mayor interés comercial y deportivo, y fortaleciendo la posición estratégica argentina como líder global en genética equina (Clarín, 2013).

Ambos ejemplos resaltan la importancia estratégica de la clonación como una herramienta poderosa no solo para preservar genética excepcional, sino también para impulsar el desarrollo competitivo del polo argentino a nivel mundial, posicionando a Argentina como un referente indiscutible en el ámbito deportivo y científico de la clonación equina.

### **1.1.3. El proceso detallado de clonación equina**

El procedimiento técnico de clonación equina es altamente sofisticado, involucrando diversas etapas especializadas e integradas (Infobae y Marcelo Regalado 2023):

1. **Extracción del ADN original:** Se obtienen células somáticas (generalmente fibroblastos) mediante biopsias realizadas en tejidos específicos como piel o médula ósea del caballo original que se desea clonar.

2. **Criopreservación celular:** Estas células son almacenadas en nitrógeno líquido a -196 °C, garantizando su viabilidad y conservación genética por largos períodos, incluso años, facilitando su posterior utilización.
3. **Obtención de ovocitos:** Se recuperan ovocitos maduros mediante aspiración folicular guiada por ultrasonido en yeguas donantes vivas o inmediatamente después del sacrificio de ejemplares destinados a este fin, asegurando condiciones óptimas de calidad ovocitaria.
4. **Enucleación del ovocito:** A través de microcirugía, se extrae cuidadosamente el núcleo del ovocito donante, eliminando así su material genético original y preparándolo para recibir el nuevo núcleo.
5. **Transferencia nuclear:** Se introduce el núcleo de una célula somática previamente cultivada en el ovocito enucleado, utilizando técnicas avanzadas de micromanipulación bajo microscopía especializada.
6. **Activación embrionaria:** Se emplean impulsos eléctricos y estímulos químicos específicos para iniciar artificialmente la división celular, emulando el proceso natural de fecundación.
7. **Cultivo embrionario en laboratorio:** Durante 5 a 8 días, los embriones clonados son cultivados en medios nutritivos especiales bajo condiciones estrictamente controladas, evaluándose continuamente su desarrollo y viabilidad embrionaria.
8. **Transferencia embrionaria:** Los embriones viables en etapas tempranas (mórula o blastocisto) son cuidadosamente transferidos al útero de yeguas receptoras sincronizadas hormonalmente y seleccionadas meticulosamente para optimizar las probabilidades de implantación exitosa.
9. **Gestación y nacimiento:** Finalmente, las receptoras llevan adelante la gestación durante aproximadamente 11 meses, culminando en el nacimiento de los potrillos clonados, genéticamente idénticos al ejemplar original.

Este procedimiento demanda múltiples transferencias embrionarias por cada clon solicitado, debido a la inherente complejidad biológica y a las tasas variables de éxito embrionario, diferenciándose sustancialmente de otros métodos reproductivos por su complejidad técnica y logística (ver **Apéndice A. El proceso de clonación equina: Pasos para clonar un caballo** para representación gráfica)

#### **1.1.4. Disponibilidad limitada de yeguas receptoras: Factor limitante clave en el proceso de clonación equina**

La disponibilidad y calidad de las yeguas receptoras representan un factor decisivo que afecta directamente el éxito reproductivo y la escalabilidad de cualquier programa de clonación equina. La selección de estas receptoras es un proceso meticuloso que involucra una serie de criterios clínicos y reproductivos estrictos, lo cual hace que el número de candidatas adecuadas sea inherentemente limitado. Este factor, frecuentemente subestimado, constituye un cuello de botella significativo para los centros de clonación, limitando su capacidad operativa y su potencial de crecimiento comercial.

#### **1.1.5. Requisitos fisiológicos y reproductivos de las yeguas receptoras**

Cada embrión generado mediante clonación requiere ser implantado en una yegua receptora que cumpla con características específicas de salud y fertilidad para llevar la gestación a término. Las condiciones fisiológicas necesarias incluyen un estado general saludable, ausencia de patologías uterinas o endócrinas, y una adecuada capacidad reproductiva demostrada. Además, estas yeguas deben ser evaluadas exhaustivamente a través de pruebas diagnósticas tales como ecografías ginecológicas, exámenes hormonales, y análisis de sangre, para asegurar que su condición corporal, nutricional y hormonal sea idónea para soportar y mantener una gestación exitosa (Hinrichs, 2005).

Por otro lado, es fundamental que las receptoras estén perfectamente sincronizadas hormonalmente con la etapa de desarrollo del embrión al momento de la transferencia. Esta sincronización se logra mediante protocolos de manejo reproductivo específicos, que incluyen el uso de tratamientos hormonales, monitoreo ecográfico constante y procedimientos veterinarios precisos para asegurar que el útero de la receptora se encuentre en un estado óptimo de receptividad endometrial (Galli et al., 2003).

#### **1.1.6. Limitaciones cuantitativas y temporales**

El centro de clonación en estudio cuenta actualmente con una capacidad operativa aproximada de 1.000 yeguas receptoras por temporada reproductiva distribuidas en dos centros de transferencia. Esta cantidad no solo refleja la disponibilidad física de animales adecuados, sino también la infraestructura necesaria para alojarlas, alimentarlas, monitorearlas, y realizar los procedimientos reproductivos requeridos. La temporada reproductiva efectiva se concentra habitualmente entre los meses de agosto y diciembre, debido a las condiciones climáticas y al ciclo reproductivo natural del caballo, lo que implica una ventana temporal relativamente estrecha para realizar todas las transferencias embrionarias planeadas.

Esta estacionalidad implica restricciones adicionales. No es posible realizar pruebas de eficiencia reproductiva fuera de esta ventana temporal, ya que todas las transferencias embrionarias deben ejecutarse simultáneamente, aumentando significativamente la complejidad logística. En consecuencia, una planificación operativa extremadamente detallada y eficiente es esencial para evitar cuellos de botella adicionales o subutilización de recursos críticos durante estos meses de alta demanda.

#### **1.1.7. Complejidades operativas adicionales y manejo de lotes embrionarios**

El proceso operativo está también condicionado por las particularidades propias de la producción de embriones clonados. En términos generales, cada maniobra de laboratorio produce alrededor de 12 embriones pertenecientes a una misma línea genética. Este enfoque permite garantizar una calidad homogénea dentro del lote embrionario; sin embargo, con frecuencia da lugar a una **sobreproducción** de embriones de una misma línea, lo que puede derivar en desafíos operativos y comerciales relacionados con su aprovechamiento.

La gestión operativa del laboratorio implica una coordinación muy precisa entre la generación y evaluación de embriones en laboratorio, la selección y preparación hormonal de receptoras, y la logística asociada a los traslados y alojamientos requeridos. Cualquier desajuste en estos procesos puede resultar en una pérdida significativa de embriones o en una subutilización de recursos valiosos, generando altos costos de oportunidad para el centro.

#### **1.1.8. Impacto comercial y económico de la escasez de receptoras**

La creciente demanda global por servicios de clonación equina ha exacerbado significativamente la limitación en la disponibilidad de yeguas receptoras. Dado que clientes internacionales suelen solicitar volúmenes considerables de clones, la presión operativa sobre los centros es elevada. Cuando la demanda supera la capacidad operativa, estos centros se ven obligados a rechazar contratos o colocarlos en listas de espera, lo cual impacta negativamente en sus oportunidades comerciales, satisfacción del cliente y, finalmente, en sus ingresos potenciales.

Por otro lado, la asignación subóptima de receptoras, como consecuencia de una planificación insuficientemente ajustada o excesivamente conservadora (asignando más embriones por clon de los estrictamente necesarios en busca de un mayor cumplimiento de contratos), reduce significativamente la eficiencia global del negocio. Cada yegua receptora innecesariamente asignada a embriones con baja probabilidad de éxito reproductivo representa una oportunidad perdida para implantar embriones más prometedores, disminuyendo la rentabilidad y comprometiendo la satisfacción y fidelidad de clientes estratégicos.

#### **1.1.9. Aspectos económicos y rentabilidad del proceso**

Desde una perspectiva económica, el proceso de clonación equina representa una actividad de alta especialización y capital intensivo, en la que los costos deben ser cuidadosamente gestionados para alcanzar niveles óptimos de rentabilidad. En el centro de clonación analizado, el costo directo por cada transferencia embrionaria representa aproximadamente un 4,5% del valor de venta de un clon, incluyendo gastos asociados a la hormonización, sincronización, monitoreo veterinario, uso de instalaciones, personal capacitado y manejo de la yegua receptora.

Cuando se consideran todos los componentes del proceso —incluidos costos de laboratorio, insumos descartables, cultivos celulares, medios de cultivo, equipamiento de alta tecnología, mano de obra especializada, logística de ovocitos y seguimiento clínico hasta el nacimiento—, el costo total por clon entregado asciende a poco más de la mitad del precio de venta. Esta cifra incluye tanto los costos directos variables como los indirectos, calculados sobre una estructura de producción que proyecta una determinada escala operativa anual.

En cuanto a la política comercial, el precio de venta estipulado se divide en un anticipo del 50% al momento de la firma y un 50% restante exigible únicamente tras el nacimiento y entrega efectiva del potrillo clonado. **Si la entrega no se concreta, el centro debe reintegrar el anticipo, asumiendo los costos de los eventuales fallos reproductivos.** Esta modalidad, beneficiosa para el cliente, obliga al centro a operar con alta eficiencia técnica y una sólida planificación financiera para evitar tensiones de liquidez.

Uno de los desafíos más complejos es la gestión de los clones excedentes. En aquellos casos en los que nacen más ejemplares de los contratados, se ofrece al cliente la opción de adquirirlos con importantes descuentos: los dos primeros a mitad de precio y los siguientes a un cuarto del valor original. Esta medida busca mitigar el impacto económico de los nacimientos no requeridos, aunque frecuentemente genera tensiones comerciales debido a que muchos clientes no desean ni están en condiciones de adquirir más ejemplares. En algunos casos, el centro termina cediendo estos potrillos sin cargo, asumiendo íntegramente su costo, sin posibilidad de revenderlos, dado que la propiedad genética pertenece al cliente original.

Desde el punto de vista financiero, el margen neto de la compañía se ubica en torno al 41,5%. Sin embargo, este margen es altamente sensible a fluctuaciones en la eficiencia de las transferencias, en la tasa de nacimientos viables y en la proporción de contratos que se cumplen

sin generar excedentes. Un solo ciclo reproductivo fallido o la generación de clones no deseados puede erosionar apreciablemente la rentabilidad de toda una temporada.

Por estas razones, es clave desarrollar estrategias que maximicen la eficiencia económica sin afectar la calidad técnica ni la satisfacción del cliente. La respuesta no pasa exclusivamente por incrementar precios, sino por optimizar el uso de los recursos más críticos del proceso: la disponibilidad y gestión de yeguas receptoras, la calidad embrionaria y la capacidad predictiva sobre los resultados reproductivos, utilizando para ello modelos sustentados en datos históricos robustos y confiables.

**Nota:** Todos los valores numéricos han sido expresados de forma relativa con el fin de resguardar la confidencialidad de la información económica analizada.

#### **1.1.10. Necesidad estratégica de optimización operativa**

Frente a los desafíos anteriormente expuestos, surge la necesidad de implementar un enfoque estratégico que permita optimizar el uso de recursos reproductivos limitados y mejorar los indicadores de rentabilidad del proceso de clonación. Actualmente, el modelo operativo del centro establece una asignación fija de embriones por clon solicitado, comúnmente establecida en torno a siete embriones. Esta cifra ha sido históricamente adoptada como promedio empírico para maximizar las probabilidades de obtener al menos un nacimiento viable por contrato. No obstante, dicha cifra no considera variaciones relevantes en la calidad embrionaria ni en la eficiencia reproductiva entre distintas líneas genéticas.

La lógica de asignación fija, si bien práctica desde una perspectiva operativa, puede derivar en importantes ineficiencias. Por ejemplo, líneas genéticas con alta eficiencia reproductiva podrían requerir un menor número de embriones para lograr el mismo resultado, mientras que otras más inestables podrían beneficiarse de un enfoque más robusto. Asimismo, asignar más embriones de los necesarios a una sola línea implica el uso innecesario de yeguas receptoras, restringiendo la capacidad del centro para atender otros contratos y reduciendo la diversificación de su cartera genética.

En este contexto, la presente investigación propone diseñar e implementar un modelo analítico de optimización basado en datos históricos de eficiencia por línea genética, calidad embrionaria observable y tasas de éxito reproductivo. El objetivo no es establecer una asignación adaptativa en tiempo real (imposible con la tecnología existente ya que no hay forma de conocer a priori la predisposición genética de un ejemplar a ser clonado), sino identificar un nuevo número fijo óptimo de transferencias por clon solicitado, más eficiente que el actual, y definido a partir de características medibles del embrión. Al reducir el número de transferencias necesarias sin afectar la tasa de éxito, se busca liberar capacidad operativa, minimizar la aparición de clones excedentes y, fundamentalmente, favorecer una mayor diversificación genética, lo que se traduce en una gestión más resiliente del riesgo biológico asociado a cada línea.

Este enfoque de optimización no solo tiene implicancias operativas directas, sino que también representa una herramienta poderosa para la toma de decisiones estratégicas en el largo plazo. La capacidad del centro para maximizar su output sin incrementar proporcionalmente sus costos fijos ni su infraestructura física constituye una ventaja competitiva significativa en un mercado global de alto valor, altamente especializado y cada vez más exigente.

### **1.2. Problema**

#### **1.2.1. Limitaciones del modelo actual de transferencia embrionaria**

El modelo operativo actualmente implementado por el centro de clonación equina se basa en la asignación uniforme de siete embriones por cada clon solicitado. Esta cifra se fundamenta en un promedio histórico de resultados aceptables, pero carece de un sustento metodológico optimizado que contemple la variabilidad interindividual de los embriones y las particularidades asociadas a cada línea genética. Esta generalización representa una limitación relevante, ya que se prescinde de un análisis estadístico diferenciado por subgrupos, lo que restringe el potencial de mejora del desempeño reproductivo (Vélez & Hinrichs, 2011).

Una de las principales restricciones del esquema actual es su escasa adaptabilidad. La utilización de un número fijo de transferencias por contrato no considera diferencias fundamentales entre embriones con distintas calidades morfológicas, grados de desarrollo ni condiciones de conservación, como el uso de embriones frescos frente a vitrificados. Tales diferencias han demostrado influir significativamente en las tasas de implantación y desarrollo embrionario, lo cual torna ineficiente el uso de recursos críticos como las yeguas receptoras, cuya disponibilidad está limitada por factores estacionales, sanitarios y logísticos (Galli et al., 2013).

Adicionalmente, la adopción de una estructura fija no capitaliza los avances científicos y la riqueza de información acumulada por el laboratorio. A lo largo de su trayectoria, el centro ha generado una base de datos sustancial que incluye registros detallados sobre eficiencia reproductiva, tasas de éxito por línea genética y características fenotípicas de los embriones transferidos. La falta de aprovechamiento analítico de esta información para ajustar los parámetros operativos representa una pérdida estratégica en términos de eficiencia, reducción de costos y mejora de la rentabilidad global del proceso (Losinno, 2018).

Por otra parte, la rigidez del modelo limita la capacidad del centro para adaptarse proactivamente ante variaciones en la demanda de servicios, restricciones logísticas o desafíos de competitividad. En situaciones de alta demanda internacional o escasez de yeguas receptoras disponibles, la imposibilidad de ajustar dinámicamente el número de transferencias por contrato ha derivado en la necesidad de rechazar solicitudes o establecer listas de espera, afectando la capacidad de respuesta del centro sin comprometer sus estándares de calidad (Carrillo Sánchez, 2021).

En este contexto, si bien la asignación de siete transferencias ha funcionado como referencia operativa durante ciclos previos, su aplicación indiscriminada y desvinculada de las condiciones reales constituye una limitación estructural del modelo. Esta debilidad metodológica compromete tanto la eficiencia reproductiva como el potencial de escalamiento y diferenciación del centro en un entorno internacional cada vez más competitivo y orientado a la personalización y optimización de recursos.

### **1.2.2. Uso fijo del número de embriones transferidos por clon**

Si bien el uso estandarizado de un número fijo de embriones por clon —comúnmente establecido en siete transferencias— genera una serie de limitaciones ya expuestas, es importante destacar que esta práctica también responde a una necesidad operativa concreta. La razón por la cual se sostiene un número fijo, y no variable por contrato, radica en la imposibilidad técnica y biológica de conocer a priori el comportamiento reproductivo específico de cada línea genética. La propensión de una línea celular a generar nacimientos exitosos solo puede evaluarse retrospectivamente, una vez acumulada suficiente experiencia empírica con esa línea. Por tanto, no existe actualmente una forma fiable de predecir la eficiencia reproductiva de una línea antes de ejecutar el contrato de clonación.

El desafío, entonces, no radica en eliminar el modelo de número fijo, sino en diseñar un modelo robusto de decisión que permita establecer cuál es ese número ideal en base a variables conocidas y observables, y que al mismo tiempo sea suficientemente conservador como para cubrir el riesgo reproductivo que no puede preverse debido a la imposibilidad de anticipar el comportamiento específico de cada línea genética. Este enfoque preserva la lógica operativa actual pero introduce un criterio técnico que permite incrementar sustancialmente la eficiencia y minimizar los costos y riesgos derivados del exceso o defecto en la cantidad de embriones transferidos.

El uso estandarizado de un número fijo de embriones por clon, como es el caso del modelo de siete transferencias, genera una serie de problemáticas operativas, económicas y comerciales que impactan negativamente en la eficiencia global del proceso de clonación equina. A continuación, se analizan en profundidad los principales problemas derivados de esta práctica.

1. **Sobreasignación de recursos reproductivos:** La aplicación de un criterio único para todos los casos implica que, en situaciones donde bastarían menos embriones para lograr un nacimiento exitoso, se continúan utilizando los siete estipulados por protocolo. Esto lleva a una sobreutilización de yeguas receptoras, que son recursos limitados, costosos y logísticamente exigentes. Cada receptora adicional utilizada sin necesidad representa una unidad menos disponible para otros contratos o líneas celulares, reduciendo así la capacidad operativa del centro y dificultando su expansión.
2. **Desaprovechamiento de la calidad embrionaria:** No todos los embriones presentan las mismas características morfológicas ni el mismo grado de desarrollo al momento de la transferencia. El modelo actual ignora estas diferencias al aplicar la misma cantidad de embriones sin considerar su calidad relativa. Esto significa que se podrían estar transfiriendo embriones de alta calidad junto a otros de baja calidad sin realizar una optimización basada en esas observaciones, lo cual reduce la eficacia del protocolo y eleva los costos por intento reproductivo fallido. Hoy el laboratorio registra la información observable pero no hace uso de ella.
3. **Costos de oportunidad elevados:** Dado que la cantidad de yeguas receptoras es finita por temporada, cada yegua asignada a un embrión de baja probabilidad de éxito impide su utilización para otro embrión potencialmente más prometedor. Esta asignación subóptima representa un costo de oportunidad importante, especialmente en escenarios de alta demanda o de interés estratégico en determinadas líneas genéticas. En otras palabras, el modelo actual sacrifica la posibilidad de captar nuevos clientes o diversificar líneas celulares en favor de un protocolo genérico que no diferencia contextos ni optimiza recursos.
4. **Generación de clones excedentes no solicitados:** La sobreasignación intencional de embriones como mecanismo de cobertura ante fallos implica un riesgo creciente de que nazcan más potrillos de los requeridos por contrato. Estos clones excedentes, aunque biológicamente viables, suelen representar un problema financiero y comercial. En general, se ofrecen al cliente a un precio reducido —50% para los dos primeros excedentes, y 25% para los dos subsiguientes—, lo que reduce el margen de ganancia y puede incluso derivar en pérdidas si el cliente no desea adquirirlos. Además, pueden surgir dificultades logísticas en la colocación de esos animales no previstos, generando tensiones comerciales y sobrecargas operativas.

Estos problemas ponen de manifiesto la necesidad de revisar el modelo actual y avanzar hacia una lógica de asignación más racional, que mantenga un número fijo de transferencias por clon —por motivos operativos y contractuales— pero que ajuste ese número a partir de una metodología analítica basada en evidencia, capaz de maximizar las tasas de éxito y minimizar los costos y riesgos derivados del exceso o defecto de embriones transferidos.

### **1.3. Objetivo**

#### **1.3.1. Pregunta de investigación**

A partir del problema planteado y considerando la relevancia estratégica y económica que tiene para el centro de clonación la optimización en la asignación de embriones clonados, esta investigación se enfoca en dar respuesta a la siguiente pregunta central:

**¿Cómo optimizar el número de embriones transferidos por clon minimizando riesgos y maximizando eficiencia?**

#### **1.3.2. Objetivo general**

Diseñar y validar un modelo analítico, basado en evidencia empírica e información histórica generada por el centro de clonación, que permita establecer un número óptimo y fijo de embriones a transferir por clon solicitado. Este modelo incorporará como insumos clave las características observables del embrión —como su calidad morfológica, día de desarrollo y estado de conservación— a fin de estimar su probabilidad de éxito y ajustar el protocolo de transferencia a un criterio estadísticamente fundamentado. La intención no es abandonar el esquema de número fijo, necesario por razones operativas, sino optimizarlo en base a datos objetivos que permitan reducir el uso innecesario de recursos críticos como las yeguas receptoras.

Además, el modelo incluirá un enfoque de diversificación del riesgo biológico asociado a la variabilidad genética entre líneas clonadas, las cuales no pueden predecirse con certeza a priori. Al contemplar estas diferencias en la estrategia operativa general —sin modificar el número por contrato, pero sí a través de una optimización agregada— se busca mejorar la eficiencia reproductiva del sistema en su conjunto. La finalidad última es maximizar el beneficio económico del centro, elevar su capacidad de gestión técnica y financiera, y mejorar la experiencia y satisfacción de los clientes, sin comprometer la tasa de nacimientos ni generar excedentes que afecten la rentabilidad del negocio.

#### **1.3.3. Objetivos específicos**

##### **1. Identificación de variables determinantes del éxito reproductivo**

Llevar a cabo un análisis exploratorio y estadístico en profundidad sobre el conjunto de datos históricos generados por el centro de clonación con el fin de identificar qué características observables del embrión están más fuertemente asociadas al éxito reproductivo. Este conjunto de datos incluye variables como la calidad morfológica (clasificación G), el estado de conservación (fresco o vitrificado), el día de desarrollo embrionario (por ejemplo, 6, 7 u 8), y otras variables clínicas y técnicas disponibles.

El análisis se basará principalmente en enfoques de machine learning supervisado, utilizando algoritmos como árboles de decisión, random forest y XGBoost para predecir la probabilidad de éxito reproductivo a partir de las variables mencionadas. Estos modelos permiten no solo

evaluar la importancia relativa de cada predictor, sino también capturar relaciones no lineales y posibles interacciones entre variables que podrían pasar desapercibidas en modelos estadísticos tradicionales.

La validación cruzada y los conjuntos de entrenamiento/test serán utilizados para evaluar el rendimiento de los modelos, priorizando métricas como la precisión, la sensibilidad y el área bajo la curva ROC (AUC). Este enfoque técnico permitirá establecer criterios de decisión confiables para definir la combinación de características óptima que mejoren las probabilidades de éxito de un embrión.

De esta forma, se busca pasar de una lógica empírica y homogénea a un enfoque basado en datos y técnicas de machine learning, que optimice el uso de recursos críticos como las yeguas receptoras, mejore la tasa de éxito por transferencia y minimice los costos asociados a fallos reproductivos o generación de clones excedentes.

## **2. Evaluar el impacto económico-operativo del modelo optimizado y su potencial para ampliar la capacidad operativa**

Simular diferentes escenarios de reducción en el número de transferencias por clon, utilizando como base el modelo optimizado propuesto en esta tesis y comparando el escenario actual (7 transferencias por contrato) con alternativas como 5, 6 u 8 transferencias. Estas simulaciones serán realizadas mediante simulaciones de Montecarlo, integrando parámetros como las tasas de éxito históricas, los costos por transferencia, los costos por embrión, y los ingresos percibidos por cada clon nacido y entregado.

El análisis económico-operativo incluirá indicadores clave como la tasa de nacimientos exitosos, la eficiencia en el uso de yeguas receptoras, el costo total por clon entregado, la rentabilidad neta por contrato, y un índice agregado de beneficio económico denominado Commercial Benefit Ratio (CBR). También se contemplará el impacto de los clones excedentes sobre la estructura de ingresos, considerando su efecto en la rentabilidad global del proceso y en la relación con el cliente.

Además, se evaluará cómo la reducción del número de transferencias afecta la capacidad operativa del centro para escalar su volumen de contratos atendidos dentro de los límites físicos y temporales existentes. Esto permitirá estimar cuántos contratos adicionales podrían ser gestionados eficientemente por temporada sin necesidad de ampliar la infraestructura actual, maximizando así la utilización del cuello de botella operativo: las yeguas receptoras.

En conjunto, esta evaluación permitirá validar la viabilidad técnica y económica del modelo propuesto, evidenciando cómo una optimización en el número de transferencias —aun manteniendo su carácter fijo— puede generar beneficios significativos en términos de eficiencia, rentabilidad, y capacidad de expansión del centro.

### **1.3.4. Justificación y resultados esperados**

A través del logro de estos objetivos, se espera aportar una herramienta metodológica robusta, técnicamente fundamentada y operacionalmente viable para mejorar la gestión del proceso de clonación equina, integrando ciencia de datos, biotecnología reproductiva y criterios de eficiencia económica. La construcción de un modelo analítico permitirá al centro de clonación abandonar la lógica empírica generalista para avanzar hacia una lógica de gestión optimizada, basada en la evidencia acumulada a lo largo de miles de transferencias embrionarias.

Esta propuesta permitirá al centro:

- Aumentar su capacidad de toma de decisiones basadas en datos (data-driven decision making), institucionalizando procesos analíticos que integren el conocimiento técnico del laboratorio con algoritmos de predicción y optimización.
- Optimizar el uso de recursos críticos como las yeguas receptoras, que constituyen el principal cuello de botella operativo. Una mejor asignación del recurso permitirá maximizar la tasa de nacimientos por ciclo reproductivo y liberar capacidad para atender una mayor diversidad de líneas genéticas.
- Reducir la aparición de clones excedentes no deseados, lo cual no solo impacta positivamente en la eficiencia económica, sino también en la relación con los clientes, quienes valoran una ejecución precisa y ajustada a lo pactado contractualmente.
- Ampliar su volumen de operaciones sin incurrir en costos estructurales adicionales, utilizando mejor la infraestructura existente y mejorando la tasa de contratos exitosamente gestionados por temporada.
- Elevar su competitividad global y su posicionamiento como líder en biotecnología equina aplicada al polo de alto rendimiento, consolidando su reputación como centro de excelencia técnica y operativa, capaz de combinar vanguardia científica con resultados reproductivos y comerciales consistentes.

De esta manera, se contribuye no solo a mejorar la eficiencia técnica del proceso reproductivo, sino también a potenciar la sostenibilidad económica, la escalabilidad del modelo operativo y la proyección internacional de uno de los centros de clonación más importantes del mundo.

## **2. Marco teórico**

### **2.1. Fundamentos de la clonación equina y transferencia embrionaria**

La clonación equina es una biotecnología reproductiva avanzada basada en la técnica de transferencia nuclear somática (SCNT, por sus siglas en inglés), mediante la cual se genera un embrión genéticamente idéntico al individuo donante. Desde su implementación en equinos a comienzos del siglo XXI, esta tecnología ha posibilitado la replicación de animales de elevado valor deportivo y genético, aunque su eficiencia reproductiva continúa siendo inferior a la de otras técnicas como la fecundación in vitro convencional (Galli et al., 2013; Vélez & Hinrichs, 2011).

Desde el punto de vista técnico, el procedimiento de clonación involucra múltiples etapas críticas: la enucleación del ovocito receptor, la transferencia del núcleo somático del donante, la activación del cigoto reconstruido, el cultivo embrionario en condiciones controladas y la posterior transferencia del embrión a una yegua receptora hormonalmente sincronizada. Cada una de estas fases presenta desafíos técnicos y biológicos específicos, lo que resulta en tasas de éxito reproductivo relativamente bajas, que en la mayoría de los casos no superan el 15% de gestaciones exitosas por embrión transferido (Galli et al., 2013).

La eficiencia del proceso clonal depende no solo de la viabilidad intrínseca del embrión, sino también de variables asociadas al origen y calidad del ovocito, el estado fisiológico y reproductivo de la receptora, las condiciones climáticas durante la gestación y los parámetros del cultivo embrionario. Estudios previos han demostrado que factores como el estadio de

desarrollo del embrión al momento de la transferencia (por ejemplo, mórula frente a blastocisto), así como el uso de embriones frescos en lugar de vitrificados, pueden tener efectos significativos sobre la tasa de implantación y el mantenimiento de la gestación (Galli et al., 2013; Grady & Hinrichs, 2016).

Del mismo modo, la calidad morfológica del embrión ha sido reconocida como un predictor confiable de éxito reproductivo. Sin embargo, a pesar de la evidencia disponible, en muchos programas comerciales de clonación se continúa utilizando un criterio homogéneo de asignación embrionaria, sin distinguir entre embriones de diferente potencial reproductivo observable (Carrillo Sánchez, 2021; Galli et al., 2013).

Esta discrepancia entre el conocimiento científico acumulado y su aplicación en la práctica clínica evidencia la necesidad de desarrollar modelos analíticos que integren variables observables en la toma de decisiones reproductivas. El uso de técnicas estadísticas y predictivas, basadas en bases de datos históricas del propio laboratorio, permitiría optimizar la asignación de recursos críticos —como las yeguas receptoras— cuya disponibilidad es limitada y no fácilmente escalable (Losinno, 2018; Carrillo Sánchez, 2021).

En síntesis, los fundamentos técnicos y biológicos de la clonación equina y la transferencia embrionaria ponen de manifiesto tanto el potencial estratégico de estas biotecnologías como sus restricciones operativas. Esta investigación propone reducir dicha brecha mediante la formalización de un enfoque basado en evidencia empírica, con el objetivo de transformar el conocimiento técnico en decisiones clínicas más eficaces, adaptativas y económicamente sustentables.

## **2.2. Eficiencia reproductiva y calidad embrionaria**

Desde la perspectiva de la ciencia de datos aplicada a la biotecnología reproductiva, el análisis sistemático de la eficiencia reproductiva y de la calidad embrionaria constituye un eje fundamental para el desarrollo de modelos predictivos robustos y estrategias de optimización operativa. En términos generales, la eficiencia reproductiva puede definirse como la proporción entre el número de embriones transferidos y los nacimientos exitosos, mientras que la calidad embrionaria hace referencia al conjunto de atributos morfológicos y temporales observables que permiten inferir su viabilidad, potencial de implantación y desarrollo (Galli et al., 2013).

En el contexto operativo del centro de clonación analizado, se registran de manera sistemática diversas variables con potencial predictivo sobre el desenlace reproductivo. Entre las más relevantes se encuentran la calidad morfológica embrionaria (habitualmente clasificada en categorías como G1, G2, etc.), el día de evolución al momento de la transferencia (por ejemplo, día 6 versus día 7), el estado de conservación (fresco o vitrificado), el número de células visibles, y el linaje genético del ejemplar clonado. Estas variables conforman un espacio de datos multidimensional, apto para la aplicación de técnicas de aprendizaje automático supervisado orientadas a la predicción del éxito gestacional.

Investigaciones previas han establecido correlaciones significativas entre variables como la calidad morfológica y el estadio de desarrollo embrionario y las tasas de implantación y nacimiento (Grady & Hinrichs, 2016; Galli et al., 2013). No obstante, la coexistencia de múltiples factores interdependientes y la presencia de relaciones no lineales entre ellos requieren metodologías analíticas más sofisticadas que superen los modelos estadísticos tradicionales.

En esta investigación se implementan modelos de clasificación como árboles de decisión, Random Forest y XGBoost, con el objetivo de predecir la probabilidad de éxito de cada embrión transferido. Estas técnicas permiten estimar la importancia relativa de cada variable explicativa, así como capturar interacciones complejas entre atributos que no serían evidentes mediante aproximaciones lineales. Los algoritmos de ensamblado como Random Forest y XGBoost han mostrado una capacidad predictiva destacada en entornos biomédicos, especialmente en escenarios caracterizados por datos parcialmente estructurados, ruidosos o con distribución desequilibrada de clases.

El propósito último de este enfoque es transformar el acervo descriptivo acumulado por el laboratorio —históricamente utilizado con fines clínicos y evaluativos— en herramientas prescriptivas que orienten la toma de decisiones operativas. De esta manera, el análisis de calidad embrionaria trasciende su función clasificatoria subjetiva y se convierte en una variable instrumental dentro de un marco analítico riguroso que determina, por ejemplo, el número óptimo de embriones a transferir por clon solicitado, en función de su probabilidad estimada de éxito. Esta estrategia representa un cambio de paradigma hacia una lógica de decisión basada en evidencia, que permite mejorar la asignación de recursos, reducir la incertidumbre del proceso y maximizar la eficiencia reproductiva global del sistema.

### 2.3. Optimización basada en datos: modelos analíticos y toma de decisiones

En el contexto de la biotecnología reproductiva aplicada al ámbito equino, la toma de decisiones fundamentada en datos empíricos representa una ventaja competitiva crítica para los centros de clonación. Estos datos, acumulados a lo largo de múltiples temporadas, ofrecen una oportunidad estratégica para modelar, anticipar y optimizar los resultados reproductivos, reduciendo la incertidumbre inherente al proceso y mejorando la asignación de recursos escasos como las yeguas receptoras (Galli et al., 2013).

Particularmente en la clonación, el cuello de botella operativo más relevante es la disponibilidad limitada de receptoras, cuya expansión no es fácilmente escalable por restricciones sanitarias, estacionales y económicas. Esto impone la necesidad de maximizar el rendimiento de cada transferencia embrionaria. En lugar de aplicar una cantidad fija e indiferenciada de embriones por contrato —práctica aún común en algunos protocolos comerciales—, es posible diseñar modelos analíticos que determinen, para cada clon solicitado, el número óptimo de embriones a transferir según su probabilidad estimada de éxito, integrando variables como la calidad morfológica, el estado de conservación (fresco o vitrificado) y el día de desarrollo (Grady & Hinrichs, 2016; Carrillo Sánchez, 2021).

Desde un enfoque metodológico, este tipo de optimización puede abordarse mediante **simulaciones de Montecarlo**, que permiten modelar escenarios múltiples incorporando distribuciones empíricas de eficiencia, restricciones operativas y estructuras de costos. Asimismo, puede integrarse un índice económico como el *Commercial Benefit Ratio* (CBR), que pondera ingresos potenciales, costos por transferencia y embrión, penalizaciones por excedentes y riesgo comercial, ofreciendo así un marco cuantitativo integral para evaluar decisiones (Netessine & Shumsky, 2002).

Complementariamente, pueden aplicarse **modelos de optimización basados en machine learning** —como random forest o Xgboost— para identificar configuraciones eficientes bajo múltiples factores simultáneos. Esta transformación de probabilidades en decisiones clínicas

concretas conecta el conocimiento técnico con el uso eficiente de los recursos disponibles (Aguiar, 2004).

Una ventaja clave de estos modelos es su capacidad para realizar análisis de sensibilidad y construir escenarios contrafactuales, permitiendo explorar la robustez del sistema frente a cambios en la eficiencia o la capacidad operativa. Esta flexibilidad analítica habilita ajustes dinámicos en los protocolos reproductivos y favorece la toma de decisiones adaptativa y basada en evidencia. Además, su arquitectura permite incorporar nueva información conforme se dispone de datos más recientes, facilitando su recalibración y mejora continua.

En definitiva, la optimización basada en datos articula biotecnología, ciencia de decisiones y estrategia empresarial. En el caso de la clonación equina, su aplicación permite trascender modelos homogéneos y orientarse hacia decisiones operativas más eficientes, fundamentadas y económicamente racionales.

### **2.3.1. Técnicas de aprendizaje automático aplicadas a la predicción de eficiencia reproductiva**

El avance de la ciencia de datos ha permitido trasladar al ámbito reproductivo equino herramientas analíticas que tradicionalmente se empleaban en otros sectores. Particularmente, los algoritmos de aprendizaje automático supervisado (machine learning) permiten construir modelos predictivos robustos capaces de anticipar resultados reproductivos en función de variables observables del embrión, como la calidad morfológica (clasificación G), el día de desarrollo y el estado de conservación (fresco o vitrificado).

Entre las técnicas más utilizadas en este tipo de problemas se encuentran los árboles de decisión, el random forest y el algoritmo XGBoost. Los árboles de decisión son modelos interpretables que dividen iterativamente los datos según umbrales en las variables independientes, generando reglas claras de clasificación. Su principal ventaja es la facilidad con la que los expertos pueden entender y validar las predicciones, algo fundamental en contextos clínicos o biotecnológicos (Quinlan, 1986).

El random forest, por su parte, es un ensamble de múltiples árboles de decisión contruidos sobre subconjuntos aleatorios de los datos y de las variables predictoras. Esta técnica mejora la precisión general del modelo y reduce el riesgo de sobreajuste, al tiempo que permite calcular la importancia relativa de cada variable en la predicción del éxito reproductivo (Breiman, 2001). Esta capacidad resulta especialmente valiosa en contextos donde se dispone de un conjunto limitado pero informativo de variables clínicas y operativas.

XGBoost (Extreme Gradient Boosting) es otro algoritmo de ensamble basado en árboles que ha demostrado resultados sobresalientes en competencias internacionales de predicción. Su ventaja principal radica en su eficiencia computacional y en la capacidad de manejar relaciones no lineales y efectos de interacción entre variables, lo que lo convierte en una herramienta poderosa para capturar la complejidad biológica de los procesos de clonación (Chen & Guestrin, 2016). En el marco de este estudio, el modelo XGBoost será entrenado con los datos históricos del centro de clonación para predecir la probabilidad de éxito de cada embrión transferido.

La elección de estas técnicas responde tanto a su capacidad de predicción como a su aplicabilidad práctica. A diferencia de los modelos estadísticos tradicionales —como la regresión logística— que requieren supuestos específicos sobre la distribución de los datos, los algoritmos de aprendizaje automático se adaptan mejor a contextos con relaciones complejas, no lineales

y con variables que pueden tener estructuras jerárquicas o interacciones entre sí. Además, permiten incorporar validación cruzada, métricas de precisión y análisis de sensibilidad, fortaleciendo la confianza en los resultados y la posibilidad de escalarlos operativamente.

### 2.3.2. Simulación Montecarlo en contextos biotecnológicos: aplicaciones a la clonación equina

La simulación Montecarlo es una técnica de análisis cuantitativo ampliamente utilizada en contextos de alta incertidumbre, donde la complejidad del sistema impide el uso de modelos deterministas. En el ámbito de la clonación equina, esta herramienta resulta particularmente útil para modelar la variabilidad inherente al proceso reproductivo y estimar, a partir de distribuciones probabilísticas, la probabilidad de éxito asociada a diferentes escenarios de asignación de embriones por clon solicitado.

Este tipo de simulación se basa en la generación de un número elevado de escenarios posibles mediante la repetición aleatoria de procesos que siguen una determinada distribución de probabilidad. En el caso del centro de clonación analizado, la variable principal de interés es la cantidad de clones exitosamente nacidos a partir de un determinado número de embriones transferidos.

Aplicar Montecarlo permite, por ejemplo, responder preguntas como: “¿Cuales son las probabilidades de que, transfiriendo cinco embriones con una tasa individual de éxito del 20%, se obtenga al menos un potro viable?”. Estas respuestas, que son clave para la toma de decisiones clínicas y comerciales, no pueden obtenerse mediante modelos estáticos ni reglas empíricas simples.

Uno de los mayores aportes de esta metodología es su capacidad para modelar los efectos acumulativos del riesgo reproductivo. Como la probabilidad de éxito de cada embrión se basa en sus características observables (por ejemplo, un embrión G1 vitrificado en día 7 puede tener una tasa estimada de éxito del 25%), la simulación permite calcular, en función del número de intentos, la probabilidad acumulada de éxito. Esto es especialmente relevante en el modelo de negocio del centro, donde la transferencia de siete embriones por clon actúa como un “seguro estadístico” frente a la incertidumbre. Montecarlo permite evaluar si ese número realmente ofrece el balance óptimo entre riesgo, costos y eficiencia operativa.

Además, esta técnica puede integrarse con métricas económicas como el **Commercial Benefit Ratio (CBR)**, que combina la rentabilidad por clon entregado, el costo total por ciclo y las penalidades asociadas a clones excedentes o fallas contractuales. De este modo, el análisis no solo considera la tasa de éxito biológico, sino también su impacto financiero, permitiendo optimizar el número de embriones desde una perspectiva integral.

En biotecnología animal, la simulación Montecarlo parece ser la indicada para evaluar estrategias reproductivas, optimizar protocolos de transferencia embrionaria, y estimar retornos económicos bajo distintos escenarios. Su aplicación al proceso de clonación equina se alinea con esta tendencia de analítica avanzada, posicionando al centro no solo como un referente técnico, sino también como una organización guiada por la toma de decisiones científicas y orientadas a valor.

Finalmente, esta técnica resulta particularmente compatible con la estructura de datos del centro, que cuenta con registros históricos detallados sobre la eficiencia de cada línea genética, la calidad de cada embrión, y los resultados de cada transferencia. Esta riqueza informativa

permite calibrar el modelo con un alto grado de realismo y aplicar la simulación no solo de manera retrospectiva, sino también como herramienta prospectiva para la planificación de temporadas futuras.

## **2.4. Integración conceptual: hacia un modelo predictivo y operativo de transferencia embrionaria**

La revisión conceptual realizada hasta aquí ha puesto en evidencia que el proceso de clonación equina se ubica en la intersección entre la biotecnología reproductiva, la ciencia de datos y la gestión estratégica de recursos. La calidad embrionaria, la eficiencia reproductiva, las restricciones operativas (como la disponibilidad limitada de yeguas receptoras) y las implicancias económicas del proceso, forman un sistema complejo, cargado de incertidumbre, donde la toma de decisiones debe ser cuidadosamente fundamentada.

A partir de este diagnóstico, la presente tesis propone una integración metodológica en la que la **analítica predictiva**, la **simulación Montecarlo** y los **modelos de optimización** se combinan para abordar el problema desde múltiples dimensiones. La premisa central es que, si bien no es posible conocer a priori la eficiencia reproductiva específica de cada línea genética, sí es posible predecir —con un grado aceptable de confianza— la probabilidad de éxito de un embrión a partir de sus características observables. Esta información puede, entonces, alimentar modelos que permitan establecer el número óptimo de transferencias necesarias para cumplir con el contrato de clon solicitado, minimizando el uso excesivo de yeguas y reduciendo la aparición de clones excedentes.

La combinación de enfoques analíticos avanzados permite el diseño de una estrategia reproductiva optimizada, en la cual la decisión sobre la cantidad de embriones a transferir no se fundamenta únicamente en promedios históricos, sino en un análisis integrado de tipo técnico, económico y comercial. Al incorporar métricas como el *Commercial Benefit Ratio* (CBR) dentro de modelos de simulación, se logra alinear las decisiones clínicas con los objetivos financieros del centro. Esto establece un marco de trabajo coherente, adaptado a la realidad operativa, y orientado a la sostenibilidad económica del sistema reproductivo.

## **3. Datos**

La base de esta investigación original está compuesta por 3.069 registros correspondientes a embriones clonados, generados por el centro de clonación equina bajo estudio. Estos datos abarcan cuatro temporadas reproductivas consecutivas (julio a junio): 2021/22, 2022/23, 2023/24 y la temporada 2024/25 hasta el 31 de diciembre. Se excluyeron registros de años anteriores por presentar inconsistencias en la carga y falta de homogeneidad en los criterios de clasificación.

Los registros reflejan las operaciones de dos centros de transferencia embrionaria (Centro A y Centro B) y documentan con precisión cada intento de implantación. La información original fue recolectada en hojas de cálculo independientes, organizadas por centro y temporada, lo que implicó una primera etapa crítica de consolidación y estandarización. Este proceso incluyó la unificación de nomenclaturas, corrección de errores y normalización de variables clave, mediante herramientas de procesamiento de datos en Python.

El dataset final integra variables de laboratorio (calidad embrionaria, día de evolución, estado de conservación), operativas (número de transferencia, centro, línea genética), y resultados

reproductivos (estado de preñez, nacimiento, entrega final). Esta base constituye el insumo fundamental para los análisis estadísticos, exploratorios y predictivos desarrollados en los capítulos siguientes.

### 3.1. Origen de la información y proceso de unificación

La información proviene de los registros operativos de dos centros de transferencia embrionaria que colaboran con el laboratorio de clonación. Cada procedimiento fue documentado en hojas de cálculo separadas por centro y por ciclo reproductivo, lo que generaba una fragmentación que dificultaba el análisis agregado.

Aunque el diseño general de las planillas era similar, se identificaron diferencias menores en la denominación de columnas, abreviaturas, ortografía y codificación de valores. Algunas variables críticas, como el tipo y estado del embrión o la calidad morfológica, presentaban variantes según el centro o la temporada. Por ello, se aplicó un proceso riguroso de estandarización semántica y estructural para garantizar la homogeneidad del conjunto de datos final.

Para llevar a cabo este proceso de integración, se utilizaron herramientas de procesamiento de datos en Python. La metodología aplicada incluyó cinco etapas principales:

1. **Normalización de columnas:** se estandarizaron los nombres de las columnas para evitar duplicidades y ambigüedades. Por ejemplo, se unificaron denominaciones como “FRESCO/VITRIF.” y “FRESCO/VITRIFICADO”, respetando la taxonomía técnica del laboratorio.
2. **Conversión de formatos y concatenación:** todas las hojas de cálculo fueron convertidas a formato CSV y concatenadas en una única tabla estructurada. Se creó además una variable adicional para identificar el centro de origen (A o B) y se codificaron las temporadas en un formato estándar (21/22, 22/23, etc.).
3. **Limpieza y validación de registros:** se eliminaron filas duplicadas y se identificaron registros con datos faltantes en campos clave como fecha de transferencia, día de evolución o estado de preñez. En los casos posibles, los valores se completaron contrastando con registros físicos impresos. Las observaciones no recuperables fueron eliminadas, representando menos del 0,4% del volumen de datos total.
4. **Homogeneización de categorías:** se reordenaron y estandarizaron los valores de variables cualitativas. En el caso de la calidad embrionaria, se definió una escala jerárquica donde G1 representa la calidad más alta, seguida de G1.5, G2, etc. Esta clasificación se aplicó de manera uniforme a todas las temporadas. Del mismo modo, se unificaron los criterios de clasificación de estadios embrionarios (mórula, blastocisto, blastocisto expandido) y el formato de codificación del estado de conservación.
5. **Trazabilidad y registro de transformaciones:** todas las modificaciones realizadas fueron documentadas y almacenadas en scripts versionados. Esto garantiza la reproducibilidad del proceso y permite actualizar el dataset si se incorporan nuevas temporadas o si se requiere retroceder alguna transformación.

El resultado fue un único archivo estructurado de 2.675 registros completos, en formato CSV, con variables limpias, consistentes y aptas para análisis cuantitativo. Esta base de datos constituye el insumo principal para los modelos estadísticos, predictivos y de simulación que se desarrollan en las secciones siguientes.

### 3.2. Composición del dataset y definición de variables

El conjunto consolidado de datos reúne variables de carácter operativo, biológico y reproductivo, las cuales permiten describir con alto nivel de detalle cada procedimiento de transferencia embrionaria realizado entre las temporadas 2021/22 y 2024/25. Esta estructura multivariable constituye el insumo principal para el desarrollo de esta tesis.

A continuación, se describen las principales variables incluidas en el dataset final:

- **Fecha de transferencia:** corresponde al día exacto en que se realizó la implantación del embrión en la yegua receptora. Esta variable permite agrupar las observaciones por ventana temporal, asociarlas a eventos contextuales (climáticos o sanitarios) y calcular métricas por ciclo reproductivo (fecha probable de parto).
- **Temporada reproductiva:** se codificó como una variable categórica (21/22, 22/23, 23/24 y 24/25), asignada en función del calendario de trabajo del centro. Esta clasificación facilita el análisis evolutivo de indicadores clave a lo largo de los años.
- **Atributos del embrión:**
  - **Calidad embrionaria:** variable ordinal (G1, G1.5, G2, G2.5, etc.) que representa el juicio morfológico del laboratorio sobre la viabilidad del embrión. Cuanto menor el valor, mayor la calidad estimada. Se basa en parámetros visuales como simetría, compactación celular y ausencia de fragmentación. Se clasifica así siguiendo un estándar propio del laboratorio. En las últimas temporadas se obtuvieron resultados erráticos a partir de esta clasificación, ya que embriones de calidad superior no aseguraban mejor performance en la preñez.
  - **Estadio o tamaño:** variable categórica que identifica el estadio de desarrollo al momento de la transferencia (mórula, blastocisto, blastocisto expandido, entre otros).
  - **Días de evolución:** variable numérica continua (5.5 a 8), que representa la edad del embrión en días desde su generación hasta la transferencia o vitrificación. Es un parámetro crítico para evaluar la madurez embrionaria.
  - **Estado de conservación:** variable dicotómica que distingue entre embriones frescos (transferidos sin criopreservación) y vitrificados (descongelados previo a la transferencia).
- **Resultados reproductivos:**
  - **Estado de preñez:** indicador binario para cada uno de los controles ecográficos realizados entre los días 30 y 330 post-transferencia (11 indicadores). Se considera que la preñez sigue su curso si todos son 1 o nulos. En caso de indicar 0, se interpreta como que la preñez se encontró sin vida en ese control.
  - **Eventos posteriores:** variables binarias que registran si el embrión resultó en aborto previo al momento de parto, nacimiento vivo o nacimiento muerto, muerte natural (una vez nacido) o eutanasia (por problemas de salud del potrillo). Permiten rastrear el desenlace biológico completo.
  - **Entrega:** Variable clave que indica si el potro nacido fue entregado al cliente (1) o no (0), ya sea por muerte, decisión comercial o por estar aún pendiente al momento del cierre de los datos. Un clon entregado es un clon exitoso.
- **Características operativas:**
  - **Centro de transferencia:** variable categórica (A o B) que identifica el establecimiento en el cual se realizó el procedimiento. Ambos son muy similares en cuanto a prestaciones y cuidado de las receptoras.

- **Línea celular:** variable alfanumérica que indica el origen genético del embrión, y que permite agrupar las transferencias asociadas a un mismo clon solicitado por un cliente. Por ejemplo, este campo podría ser Cuartetera y los clones entregados de esta línea serían genéticamente idénticos a ella.
- **Número de transferencia:** entero que refleja el orden transferencias sucesivas de esa línea celular. Es fundamental para calcular cuántos embriones fueron necesarios hasta lograr un potro entregado.

Esta estructura de variables habilita un análisis exhaustivo de las transferencias tanto a nivel individual como agregado, y permite estudiar relaciones entre calidad embrionaria, condiciones operativas y resultados reproductivos. Además, es la base para construir modelos de predicción y simulación destinados a optimizar la asignación de recursos en futuras temporadas.

### 3.3. Control de datos faltantes e inconsistencias

Durante el proceso de limpieza y estructuración del dataset, se realizó un control sistemático para identificar observaciones con valores ausentes, inconsistencias internas y posibles errores de carga. El análisis se centró en variables críticas para el estudio, tales como la fecha de transferencia, el día de evolución embrionaria, la calidad embrionaria y el resultado reproductivo.

Se detectó un número acotado de registros con información incompleta en campos esenciales. En la mayoría de los casos, fue posible recuperar los datos faltantes cotejando las hojas de cálculo digitales con los registros físicos del laboratorio. Este procedimiento permitió conservar la integridad de los casos sin comprometer la calidad del conjunto. En los pocos registros donde la recuperación no fue posible, las observaciones fueron descartadas. En total, se eliminaron menos del 0,4 % de los casos, sin impacto estadístico significativo sobre la muestra.

En relación con la presencia de valores atípicos (outliers), se evaluaron distribuciones y rangos esperables para las principales variables numéricas. No se observaron valores extremos anómalos en el campo "Día de evolución", ni desviaciones inusuales en la variable "Calidad". Del mismo modo, no se registraron temporadas con criterios de clasificación embrionaria sustancialmente distintos, lo cual permite asumir homogeneidad metodológica a lo largo del período 2021–2024.

Finalmente, se verificó que la escala de calidad embrionaria (G1, G1.5, G2, etc.) fue utilizada de forma consistente entre los diferentes centros y temporadas. Esta estabilidad permite realizar análisis comparativos entre ciclos reproductivos sin necesidad de realizar ajustes metodológicos o reescalados adicionales.

### 3.4. Exploración descriptiva

Con el propósito de brindar una perspectiva inicial sobre la estructura y consistencia del dataset, se incluyó la Tabla 1, que exhibe siete registros a modo ilustrativo. Esta muestra refleja las columnas principales y permite observar cómo se organizan los campos antes de proceder a los análisis estadísticos. Cabe aclarar que los valores han sido modificados ligeramente para no exponer datos confidenciales.

| FECHA      | LÍNEA CELULAR | Nº TRANSFERENCIA | Nº EMBRION | EMBRION          | CALIDAD | FRESCO O VITRIFICADO | DÍA | CENTRO | PARTO | NACIDO VIVO | EUTANASIA | ENTREGA |
|------------|---------------|------------------|------------|------------------|---------|----------------------|-----|--------|-------|-------------|-----------|---------|
| 10/12/2021 | AAA           | 357              | 1          | BLASTOCISTO EXP. | G1      | FRESCO               | 7   | A      | 1     | 1           | null      | 1       |
| 10/12/2021 | CCC           | 366              | 1          | BLASTOCISTO EXP. | G1      | VITRIFICADO          | 7   | B      | 1     | 1           | null      | 1       |

|            |     |     |   |                  |      |             |   |   |   |      |      |   |
|------------|-----|-----|---|------------------|------|-------------|---|---|---|------|------|---|
| 10/12/2021 | AAA | 358 | 2 | BLASTOCISTO EXP. | G1.5 | FRESCO      | 7 | A | 0 | null | null | 0 |
| 20/10/2021 | BBB | 110 | 1 | BLASTOCISTO      | G1   | VITRIFICADO | 6 | A | 1 | 1    | null | 1 |
| 22/10/2021 | BBB | 117 | 1 | BLASTOCISTO      | G1   | VITRIFICADO | 6 | A | 1 | 1    | null | 1 |
| 5/11/2021  | CCC | 184 | 3 | BLASTOCISTO      | G1   | VITRIFICADO | 6 | A | 1 | 1    | 1    | 0 |
| 17/11/2023 | DDD | 232 | 6 | BLASTOCISTO      | G1   | VITRIFICADO | 6 | B | 1 | 1    | null | 1 |

Tabla 1. Ilustración resumida del dataset.

El interés de ilustrar la Tabla 1 radica en la interpretación de cómo se comportan los registros. El caso del primer registro describe una transferencia embrionaria exitosa realizada el 10 de diciembre de 2021 en el Centro A. El embrión transferido, proveniente de la línea genética "AAA" (primer embrión de esta línea), era de alta calidad (G1) y en estadio avanzado (blastocisto expandido), transferido en fresco a los 7 días de desarrollo embrionario. La transferencia resultó en el nacimiento exitoso de un potro vivo, el cual posteriormente fue entregado al cliente, cumpliendo así satisfactoriamente con el contrato de clonación.

Este tipo de análisis puede replicarse para cada registro y utilizarse luego para construir modelos estadísticos, que permitan identificar los factores determinantes para una transferencia exitosa y así optimizar el proceso operativo del laboratorio.

A fin de obtener una visión más completa sobre el comportamiento del conjunto de datos, se han generado una serie de visualizaciones que permiten explorar la distribución de las variables clave. A continuación, se presentan seis gráficos que ilustran las tendencias generales observadas en el dataset:

### 3.4.1. Distribución de transferencias por temporada:

Un gráfico de barras apiladas como se observa en la Ilustración 1 refleja el total de transferencias realizadas cada ciclo anual, desglosando su resultado en “Entregadas”, “Fallidas” y “Preñeces en curso”. Este enfoque evidencia tendencias globales y variaciones año a año.

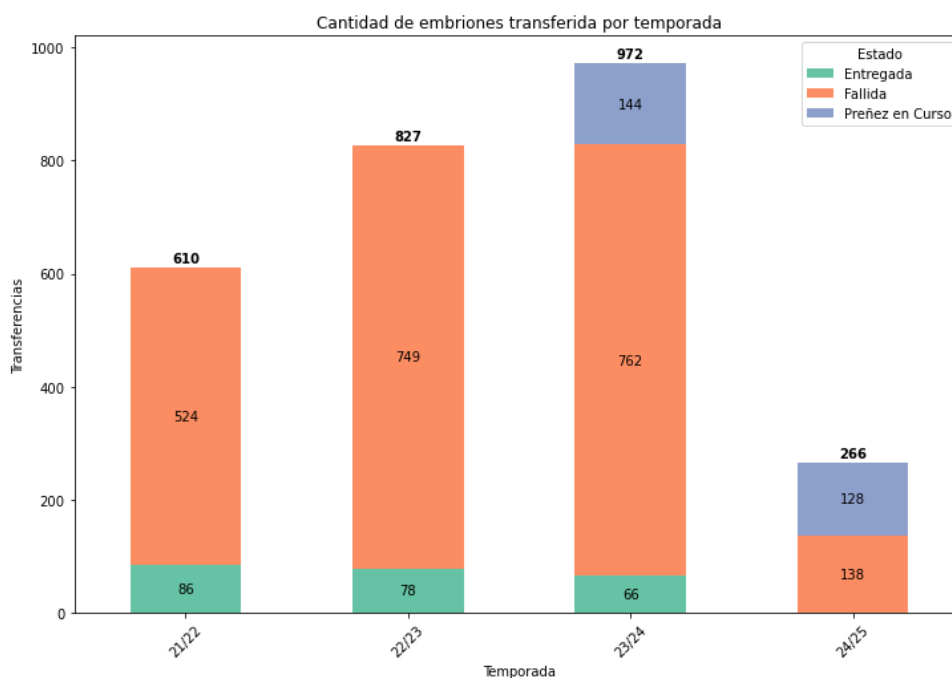


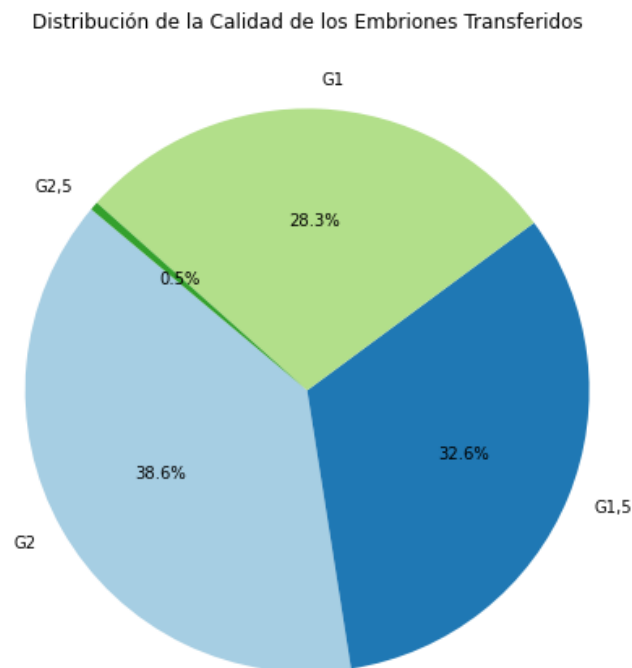
Ilustración 1 Cantidad de embriones transferida por temporada (2021-2025).

Los datos muestran que cada temporada (21/22, 22/23, 23/24, 24/25) presenta un número diferente de transferencias y resultados reproductivos (Entregadas, Fallidas y Preñeces en curso). Se observa que:

- **Temporadas 21/22:** Muestran tendencias estándar, sin grandes sobresaltos en los registros. Se considera que esta fue una temporada “normal” en cuanto a factores externos y productividad.
- **Temporada 22/23:** Se incrementó el número de transferencias, pero el éxito global disminuyó. Esta situación se asocia a factores ambientales adversos, principalmente la sequía y las altas temperaturas durante el otoño de 2023, que afectaron las gestaciones.
- **Temporada 23/24 y 24/25:** Exhibe una mayor proporción de preñeces en curso, en estado avanzado, lo que sugiere una tendencia positiva preliminar. Sin embargo, dichas preñeces todavía no se consideran "exitosas" para los modelos estadísticos (no han nacido), aunque sí se contabilizan las que resulten fallidas, incrementando así el tamaño de muestra (n) de las observaciones.

### 3.4.2. Calidad de los embriones transferidos:

El laboratorio asigna categorías cualitativas al embrión (G1, G1.5, G2, etc.). En un diagrama de pastel, como se observa en la Ilustración 2, se visualiza la proporción de cada clase de calidad, siendo G2 la categoría mayoritaria.



*Ilustración 2. Distribución de la calidad de los embriones transferidos.*

Aunque G1 dominaría en cuestiones de calidad, no se concluye de manera absoluta que la mejor calidad declarada conlleve siempre un éxito asegurado. En relación con este punto:

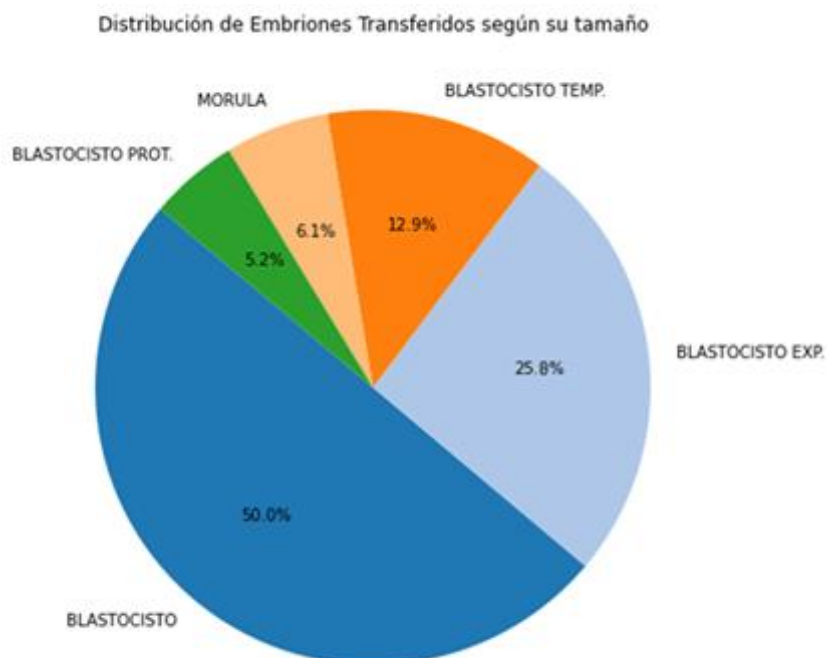
- **Incertidumbre de la correlación “Calidad-Éxito”:** si bien el indicador de calidad busca descartar embriones con desviaciones severas del estándar, no se tiene

la certeza de que la categoría G1 resulte en un nacimiento exitoso en todos los casos. Por ello, es fundamental estudiar la efectividad comparando la tasa de preñez y la entrega final por cada categoría de calidad.

- **Protocolo y decisión operativa:** la clasificación actual es parte de un protocolo interno de laboratorio que, en general, permite descartar embriones con mala proyección, maximizando el uso de receptoras. Sin embargo, el impacto real de dichas categorías en la tasa de éxito final sigue siendo parte del análisis que se desarrolla en etapas posteriores del estudio.

### 3.4.3. Distribución de embriones según estadio (desarrollo):

Un gráfico de torta, representado en la Ilustración 3, agrupa los embriones en **mórula**, **blastocisto**, **blastocisto expandido** y otras categorías. Se advierte la **preeminencia de blastocistos** en distintos estadios, lo que refleja la práctica estándar de transferir embriones en fases avanzadas para optimizar la implantación.



*Ilustración 3. Distribución de embriones transferidos según su estadio de desarrollo.*

Desde una perspectiva evolutiva del desarrollo embrionario, la secuencia progresiva comprende las etapas de **mórula**, **blastocisto temprano**, **blastocisto**, **blastocisto prototípico** y, finalmente, **blastocisto expandido**.

### 3.4.4. Estado de conservación al momento de la transferencia:

En la Ilustración 4 se compara la proporción de **embriones frescos** frente a los **criopreservados (vitrificados)** mediante un diagrama circular. Se observa que el uso de embriones vitrificados es considerable, lo que resalta la utilidad de esta técnica para administrar la disponibilidad de yeguas receptoras y mejorar la eficiencia del programa de clonación. La vitrificación permite generar o mantener embriones a lo largo del año y transferirlos en el momento de máxima receptividad uterina, evitando así la dependencia estricta del “tiempo real” de producción.

Distribución de Embriones Frescos vs Vitificados

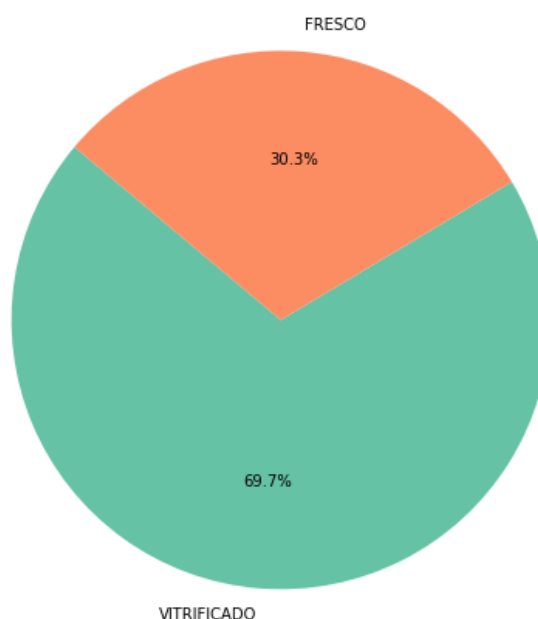


Ilustración 4. Estado de conservación de los embriones transferidos (frescos vs. vitificados).

### 3.4.5. Día de evolución del embrión previo a la transferencia:

Un gráfico de barras, como se muestra en la Ilustración 5, indica cuántas transferencias se realizaron según la edad del embrión en días (5.5, 6, 7, etc.). El pico más alto se observa en **día 7**, lo que sugiere una preferencia por realizar la transferencia en ese momento del desarrollo embrionario, asociado a su estadio evolutivo o tamaño.

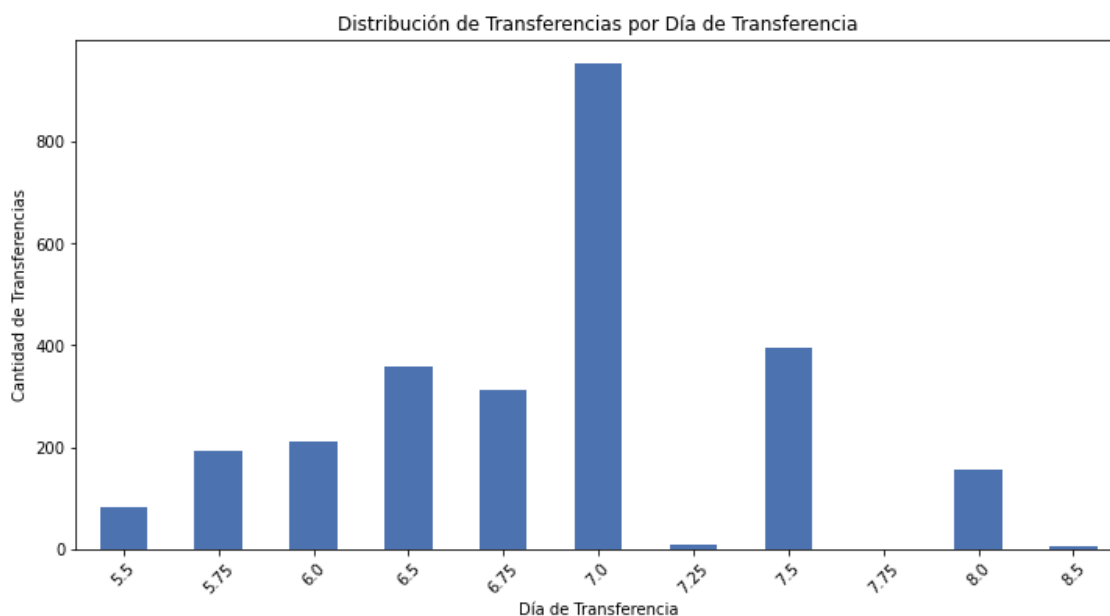


Ilustración 5. Distribución de transferencias según la edad del embrión al momento de la implantación.

Aunque la mayoría de las transferencias se concentran entre los días 6 y 7 de evolución, esto no implica que exista una ventaja operativa decisiva en optar por un día específico. En la práctica:

- **Estándar de desarrollo embrionario:** el laboratorio espera que el embrión evolucione hasta un nivel considerado óptimo antes de proceder a la transferencia o a la vitrificación. Si el embrión toma más tiempo de lo normal, podría estar asociado a problemas de desarrollo.
- **Falta de evidencia de un día “ideal”:** hasta ahora, no se considera que cierto día de evolución (por ejemplo, 7) sea intrínsecamente mejor que los demás para la implantación; más bien, se registra la variabilidad para comprender si hay fallas asociadas a un desarrollo embrionario demasiado lento o acelerado.

### 3.4.6. Distribución del promedio de transferencias por clon entregado

Una de las variables clave para entender la eficiencia genética y operativa del centro es la cantidad promedio de transferencias necesarias para lograr al menos una entrega por línea celular. La Ilustración 6 representa esta métrica bajo la forma de un histograma, en el que cada barra muestra cuántas líneas celulares alcanzan una entrega con determinada cantidad promedio de intentos. Para este análisis se excluyó la temporada 24/25, ya que al momento del corte temporal no se habían registrado entregas confirmadas, lo que hubiese sesgado los resultados al generar divisores nulos.

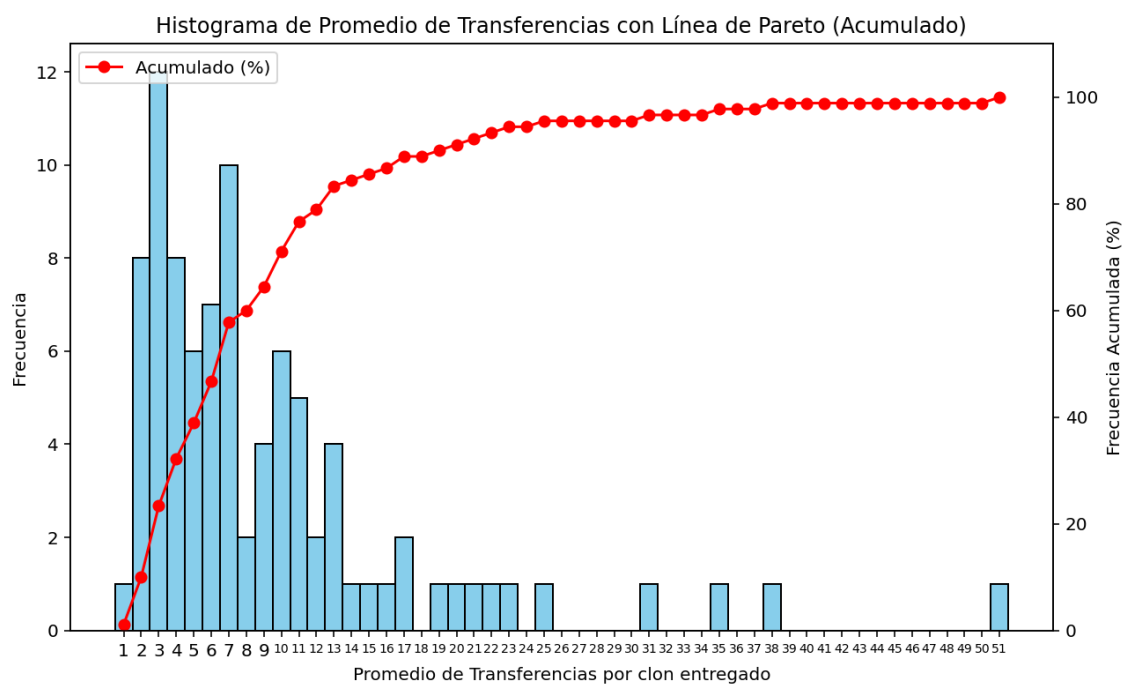


Ilustración 6. Histograma de promedio de transferencias por clon entregado (temporadas 21/22 a 23/24).

El eje x representa el promedio de transferencias embrionarias por clon nacido de cada línea celular, calculado como el cociente entre el número total de transferencias realizadas y el número total de nacimientos logrados en esa misma línea. Así, una línea con valor 7 indica que, en conjunto, se requirieron siete transferencias para obtener cada clon nacido (por ejemplo, 14 transferencias y 2 nacimientos  $\Rightarrow 14 / 2 = 7$ ). De este modo, el histograma compara la eficiencia reproductiva global entre líneas celulares y no la secuencia de intentos individuales.

Se observa que:

- Algunas líneas celulares requieren un número bajo de intentos para alcanzar una entrega, lo que sugiere una alta eficiencia reproductiva o una propensión biológica favorable. Estas líneas, aún sin haberse definido un umbral, se perfilan como más propensas a ser clonadas, ya que requieren menos recursos por unidad entregada.
- En cambio, otras líneas muestran promedios más elevados, indicando una mayor cantidad de transferencias por cada clon efectivamente entregado. Esto podría reflejar factores como ineficiencias reproductivas o características genéticas menos robustas.
- La curva de acumulado (línea de Pareto en rojo) permite visualizar qué porcentaje del total de líneas se encuentra por debajo de determinados valores promedio, lo que será relevante para definir umbrales estratégicos en etapas posteriores del análisis.

Este tipo de visualización es clave para dimensionar la dispersión en la eficiencia reproductiva por línea genética. Aunque todavía no se ha definido un umbral a partir del cual considerar una línea como “ineficiente”, este gráfico sienta las bases para dicha clasificación futura. La heterogeneidad observada subraya la necesidad de desarrollar criterios económicos óptimos, evitando así tanto la sobreasignación como la pérdida de oportunidades comerciales por falta de racionalización.

### 3.5. Cobertura temporal y validez

El lapso 2021–2024 abarca un volumen significativo de datos (2.675 registros), lo que aporta solidez estadística. No se aprecian alteraciones fundamentales en el método de recolección, ni cambios en la definición de calidades embrionarias que pudiesen sesgar el análisis. El acuerdo de confidencialidad existente entre el laboratorio y la investigación garantiza que los datos se usen únicamente con fines académicos y sin exponer información sensible o identificatoria de clientes.

## 4. Metodología

### 4.1. Modelo predictivo

En esta sección se describe el proceso metodológico seguido para analizar y predecir la probabilidad de éxito (definido como “ENTREGA”=1) de un embrión en función de sus características. El objetivo principal es identificar las variables con mayor incidencia en la entrega y, posteriormente, encontrar las combinaciones óptimas de factores (tipo de embrión, calidad, día y tratamiento fresco/vitrificado) que maximicen la probabilidad de éxito.

#### 4.1.1. Variable Objetivo (TRANSFER\_STATUS):

La variable objetivo, denominada TRANSFER\_STATUS, fue construida a partir de las columnas ENTREGA y PENDIENTE ENTREGA. La primera indica si un embrión fue efectivamente entregado (valor 1) o no (valor 0), mientras que la segunda señala si un embrión permanece en estado pendiente de entrega (1) o no (0). Para fines del modelo de clasificación, se consideró como caso positivo (TRANSFER\_STATUS = 1) tanto a los embriones ya entregados como a aquellos pendientes de entrega, dado que estos últimos representan potros listos para ser entregados y, en la práctica, siempre terminan siendo entregados, lo que permite aprovechar datos de la

temporada en curso. Por otro lado, aquellos registros que aún no tienen definido su estado (por ejemplo, porque aún no se llegó a término) fueron excluidos del análisis y marcados como NaN.

#### **4.1.2. Selección de Variables Predictoras**

El modelo incorpora como variables predictoras tanto características embrionarias como condiciones del proceso. Se incluyen: el tipo o estadio del embrión (convertido a variables dummy), la calidad morfológica (G1, G2, etc.), el estado de conservación (fresco o vitrificado) y el día de evolución embrionaria al momento de la transferencia.

#### **4.1.3. Limpieza y Tratamiento de Datos**

Para garantizar la consistencia del análisis, se eliminaron las filas con valores nulos en las variables seleccionadas. Además, las variables categóricas fueron transformadas en variables dummy, permitiendo su uso en algoritmos de machine learning.

#### **4.1.4. Balanceo de Clases**

Se identificó un desbalance en la distribución de la variable objetivo, con una proporción significativamente menor de casos positivos (TRANSFER\_STATUS = 1) respecto de los negativos. Para evitar sesgos en el entrenamiento del modelo y mejorar su capacidad de generalización, se aplicó una técnica de sobremuestreo (upsampling) sobre la clase minoritaria, replicando aleatoriamente registros hasta igualar la cantidad de observaciones de ambas clases.

#### **4.1.5. Modelado y Experimentación**

El desarrollo del modelo predictivo se abordó de manera iterativa, combinando métodos estadísticos tradicionales y algoritmos de aprendizaje automático, con el objetivo de encontrar la arquitectura con mejor desempeño para predecir la probabilidad de éxito (TRANSFER\_STATUS = 1) de un embrión clonado.

##### **4.1.5.1. Etapa exploratoria: modelos lineales**

Inicialmente, se implementaron modelos de regresión clásica: Modelo Lineal de Probabilidad, Logit y Probit. Estas herramientas permiten interpretar coeficientes y significancia estadística de las variables, facilitando una primera lectura sobre relaciones potenciales. No obstante, los resultados arrojaron baja capacidad explicativa (pseudo  $R^2 < 0.15$ ) y alta sensibilidad al desbalance de clases. Dada la complejidad no lineal del fenómeno biológico subyacente, se descartaron estos modelos en etapas posteriores del análisis.

##### **4.1.5.2. Etapa avanzada: modelos de árboles**

Se optó por algoritmos basados en árboles de decisión, debido a su capacidad para capturar relaciones no lineales e interacciones complejas entre variables.

- **Árbol de Decisión:** entrenado con profundidad limitada para evitar sobreajuste, permitió identificar la importancia relativa de variables como día embrionario y calidad morfológica. La precisión alcanzó un 61–62%, revelando potencial predictivo, pero con baja generalización.
- **Random Forest:** Con 100 y luego 200 árboles ( $n_{\text{estimators}}$ ), se ajustaron parámetros como `max_depth`, `min_samples_split` y `class_weight`. La precisión mejoró a ~63.7%, con una mayor estabilidad, aunque persistía una asimetría en el recall (mejor desempeño en la clase "Entregada").

#### 4.1.5.3. XGBoost: modelo principal

Dado el desafío de modelar fenómenos con alta incertidumbre y complejidad biológica, se empleó XGBoost como modelo final por su robustez, capacidad de ajuste fino y rendimiento superior en clasificación binaria desbalanceada.

##### Hiperparámetros clave ajustados:

- `scale_pos_weight` para contrarrestar el desbalance de clases.
- `max_depth`, `gamma` y `learning_rate` para controlar el sobreajuste.
- `n_estimators` entre 200 y 300, priorizando la estabilidad.

##### Métricas de desempeño obtenidas:

- **Precisión global:** 64–65%.
- **Recall para clase TRANSFER\_STATUS = 1:** superior al 90%, lo que lo convierte en una herramienta útil para anticipar entregas.
- **Recall para clase 0 (fallidas):** considerablemente más bajo, reflejando la mayor dificultad en identificar fallos dada la variabilidad natural no observable.

**Limitaciones biológicas:** El laboratorio estimaba a priori que un 35–40% de la variabilidad en los resultados está explicada por factores genético-naturales no incluidos en el modelo, lo que establece un techo estructural al poder predictivo de cualquier enfoque basado exclusivamente en variables observadas. Esta estimación del laboratorio concuerda con la precisión global del modelo.

#### 4.1.5.4. Optimización de Combinaciones

Con el modelo XGBoost entrenado, se procede a generar todas las combinaciones posibles de las principales variables categóricas: tipo de embrión, día de evolución, calidad morfológica y estado de conservación. Para cada combinación se calcula la probabilidad estimada de éxito.

Estas combinaciones no solo representan patrones ideales desde el punto de vista del modelo, sino que también sirven como guía operativa para ajustar los protocolos de transferencia bajo criterios empíricamente fundamentados.

**Nota:** La metodología descrita se basa en iteraciones sucesivas de prueba y error con distintos modelos, hiperparámetros y técnicas de balanceo de clase. Si bien los resultados finales no superan el 65% de precisión en la predicción global, se considera un avance significativo en la predicción de “Entregada” frente al alto grado de incertidumbre genética reportado por el laboratorio.

## 4.2. Diversificación de líneas genéticas

Una vez aplicada la primera etapa de optimización basada en variables observables el modelo metodológico propuesto avanza hacia una segunda línea de trabajo, que no busca mejorar la eficiencia individual de cada transferencia, sino **redefinir la estrategia global de asignación de recursos reproductivos**. Esta segunda etapa parte de un escenario ya optimizado en términos técnico-biológicos y propone **reducir el número fijo de transferencias por clon**, manteniéndolo constante pero en un valor menor al estándar actual, con el objetivo explícito de **diversificar las líneas genéticas atendidas por temporada** y, con ello, **reducir la exposición al riesgo biológico** de fallas específicas por línea celular.

El fundamento de esta estrategia se apoya en un concepto bien establecido en disciplinas como las finanzas: ante eventos inciertos y heterogéneos —como lo son las tasas de preñez asociadas a distintas líneas genéticas—, la diversificación actúa como un **mecanismo racional de cobertura**. Tal como plantea Markowitz (1952) en su teoría moderna de carteras, **distribuir recursos entre múltiples activos** (o en este caso, líneas genéticas) permite **reducir el riesgo agregado** sin necesidad de disminuir la expectativa de rendimiento global. Concentrar recursos (en este caso, yeguas receptoras) en un número reducido de líneas puede maximizar resultados si la eficiencia esperada se cumple, pero también puede generar pérdidas significativas cuando esas líneas presentan un comportamiento reproductivo por debajo de lo esperado. Por el contrario, **asignar un número menor de transferencias a una mayor cantidad de líneas** permite dispersar ese riesgo, reduciendo la probabilidad de que un mal desempeño individual comprometa el resultado global de la temporada.

La base empírica de esta decisión se sustenta en el análisis de la eficiencia reproductiva histórica por línea celular. A partir de los registros consolidados entre las temporadas 2021/22 y 2023/24 —excluyendo la 2024/25 por carecer aún de datos definitivos de entrega— se construyó un indicador clave: **el promedio de transferencias necesarias por línea para lograr al menos una entrega efectiva**. La Ilustración 6 muestra la distribución de esta métrica bajo la forma de un histograma con línea de Pareto acumulada. El gráfico revela una marcada heterogeneidad entre líneas: mientras algunas alcanzan entregas con muy pocas transferencias (valores promedio entre 3 y 5), otras requieren más de 10, 15 e incluso 30 intentos para lograr el mismo resultado, lo que implica un consumo desproporcionado de yeguas receptoras por unidad entregada.

#### 4.2.1. Planteo del problema y enfoque económico

Dados estos antecedentes, se propone **reducir el número fijo de transferencias** aplicadas a cada clon, evaluando escenarios por debajo del estándar actual (típicamente 7). Desde la perspectiva económica, cada transferencia supone un costo tangible —uso de la yegua receptora, insumos veterinarios, mano de obra, producción de un embrión apto— y, además, un **costo de oportunidad**: dedicar yeguas a una línea menos eficiente impide atender nuevas líneas genéticas con mejor desempeño potencial. En consecuencia, **el problema no radica únicamente en la cantidad de nacimientos logrados**, sino también en la **asignación óptima de los recursos** para maximizar la diversificación y la rentabilidad global, **evitando** tanto la sobreasignación (que genera sobre costos) como la subasignación (que podría producir pérdidas comerciales por no cumplir los contratos).

En cada escenario, se busca responder a la pregunta operativa y económica: **¿cuál es el número mínimo de transferencias por clon que permite mantener una tasa de entregas aceptable y, al mismo tiempo, reducir costos operativos o reinvertir recursos en nuevas líneas genéticas?**

Desde el punto de vista metodológico, esto implica:

- **Identificar** las líneas celulares y su eficiencia histórica (número de transferencias vs. entregas).
- **Estimar** el impacto de truncar el número de transferencias en distintos valores (p. ej., ¿qué porcentaje de líneas logran entregar en  $\leq 5$  transferencias? ¿qué pasa en 6 o 7?).
- **Cuantificar** las transferencias ahorradas y la capacidad adicional para incorporar más líneas.

- **Asociar** a cada transferencia el costo económico unitario (y eventualmente el costo fijo total, costo de embriones, etc.), evaluando así la rentabilidad neta.
- **Evaluar** el comportamiento de líneas “fallidas”, que incurren en costos sin generar ingresos.

En cada escenario, se calculan indicadores clave, tales como:

- **Porcentaje de líneas exitosas** o tasa de entregas.
- **Costo promedio por clon entregado y costo hundido** de líneas fallidas.
- **Rentabilidad neta** y ahorro económico total.
- **Commercial Benefit Ratio (CBR)** que integra la rentabilidad y la tasa de éxito en un solo valor. Este indicador es clave porque a pesar de que no se tiene una estimación de cuantos clientes se pueden perder por incumplimiento de contrato y qué representa económicamente eso, es preferible obtener el máximo beneficio a una tasa de cumplimiento lo más alta posible. Relacionar estas dos variables es fundamental para la toma de decisión.

#### 4.2.2. Implementación con Simulación Montecarlo

Para llevar a la práctica este análisis de forma realista, se implementó una **simulación estocástica de tipo Montecarlo**, que permite incorporar la incertidumbre biológica propia del proceso reproductivo. A diferencia de un análisis puramente retrospectivo, la técnica Montecarlo combina:

1. **Probabilidades empíricas de éxito** por línea, estimadas a partir de los datos históricos;
2. **Variación aleatoria** en cada intento de transferencia;
3. **Múltiples iteraciones** (1.000 por escenario) para estimar valores esperados y dispersión.

La lógica es la siguiente:

1. **Capacidad fija de yeguas receptoras:** Para un límite de  $n$  transferencias por clon, se puede atender un total de  $1000/n$  clones contratados.
2. **Costos e ingresos:** Se consideran costos fijos (p. ej., costo total de infraestructura), costos variables (número de transferencias, número de embriones producidos) e ingreso por clon exitoso. El ingreso puede contemplar, además, descuentos si se producen más clones de los contratados. En este caso, si un clon presenta éxitos excedentes, el segundo y tercer éxito generan la mitad del ingreso, el cuarto y el quinto un 25% y los restantes se entregan sin cargo.
3. **Simulación de resultados:**
  - Para cada clon, se selecciona aleatoriamente una línea genética, con su probabilidad de éxito  $p_{line}$ . Se emplea una **distribución equiprobable**.
  - Se generan  $n$  transferencias simultáneas (eventos Bernoulli), se cuenta cuántos éxitos hay, y se calcula el ingreso efectivo.
  - Se repite para tantos clones como permita la capacidad.

- Al finalizar, se obtienen valores agregados de entregas, costos totales, rentabilidad y CBR.
- Se promedian los resultados a lo largo de todas las iteraciones, lo cual ofrece estimaciones más robustas y estables.

#### 4.2.2.1. Descripción del algoritmo

El algoritmo presenta el siguiente proceso:

- **Carga de datos** y filtrado de la última temporada para evitar sesgos.
- **Cálculo de las probabilidades** por línea genética, basadas en la tasa histórica de éxito.
- **Parámetros económicos**, como el costo por transferencia, costo por embrión, ingreso por clon y capacidad de yeguas.
- **Bucle principal de simulación** que, para cada escenario (1 a 10 transferencias por clon contratado), determina la cantidad de clones a simular, realiza las transferencias, computa éxitos y calcula indicadores económicos claves.
- **Construcción de un DataFrame** (results\_df) que contiene los resultados promedio de cada escenario.
- **Generación de gráficas** para visualizar la evolución de clones exitosos, costo por clon entregado, rentabilidad y un índice de beneficio comercial.

## 5. Resultados

Esta sección presenta los principales hallazgos obtenidos a partir del enfoque metodológico desarrollado en la sección anterior. El análisis se centra en evaluar la capacidad predictiva de las variables observables sobre el éxito reproductivo, entendido como la entrega efectiva de un clon, y en explorar escenarios alternativos que permitan optimizar la asignación de recursos del centro de clonación.

Los resultados se estructuran en dos ejes complementarios:

1. **Predicción individual de éxito embrionario:** se examina el rendimiento de diferentes modelos de clasificación en la estimación de la probabilidad de entrega, identificando las variables con mayor poder explicativo y analizando las combinaciones óptimas de factores.
2. **Simulación de escenarios de reducción de transferencias:** a partir de la evidencia empírica de eficiencia histórica, se evalúan escenarios alternativos al esquema actual de siete transferencias por clon, considerando su impacto sobre la cobertura genética, la generación de excedentes y el aprovechamiento de la capacidad operativa.

Estos resultados permiten avanzar hacia una interpretación integrada de los aspectos técnicos, biológicos y operativos involucrados en la decisión sobre el número óptimo de transferencias por clon solicitado.

### 5.1. Resultados del modelo predictivo

Con el propósito de estimar la probabilidad de éxito reproductivo en función de variables observables del embrión, se implementaron distintos modelos clasificatorios que intentan

predecir la variable binaria TRANSFER\_STATUS. Tal como se explicó en la sección metodológica, esta variable asume valor 1 si el embrión fue entregado o se encuentra pendiente de entrega, y 0 si falló. Esta aproximación, validada operativamente con el laboratorio, permitió aprovechar un mayor volumen de datos provenientes de la temporada más reciente.

### 5.1.1. Desempeño general de los modelos

#### 5.1.1.1. Modelo de árbol de decisión

Para explorar la capacidad predictiva de las características embrionarias sobre el éxito reproductivo, se implementó un modelo de Árbol de Decisión ajustado sobre un conjunto de datos balanceado mediante sobremuestreo estratificado (seed = 42). El modelo fue entrenado utilizando una proporción de 80/20 para partición entrenamiento/prueba, incorporando como predictoras variables categóricas transformadas en dummies (tipo de embrión, calidad, estado de conservación) y una variable continua (día de desarrollo). La configuración final incluyó los hiperparámetros criterion='gini', max\_depth = 5, min\_samples\_split = 10 y random\_state = 42.

La evaluación del modelo sobre el conjunto de prueba arrojó una **precisión global (accuracy)** de 0,62, lo que indica que el algoritmo clasificó correctamente el 62% de los casos en la categoría correspondiente de éxito o falla reproductiva.

La **matriz de confusión** (

|      |          | Estimado por el modelo |          |
|------|----------|------------------------|----------|
|      |          | Negativo               | Positivo |
| Real | Negativo | 256                    | 220      |
|      | Positivo | 154                    | 348      |

Tabla 2) muestra un desempeño moderado en ambas clases: se identificaron correctamente 256 casos de fallos reproductivos y 348 casos de entregas exitosas, aunque se incurrió en 220 falsos positivos y 154 falsos negativos. Esto sugiere una ligera inclinación del modelo hacia la clase positiva (éxito), como se observa también en los indicadores de recall y f1-score.

|      |          | Estimado por el modelo |          |
|------|----------|------------------------|----------|
|      |          | Negativo               | Positivo |
| Real | Negativo | 256                    | 220      |
|      | Positivo | 154                    | 348      |

Tabla 2. Matriz de confusión del modelo de Árbol de Decisión.

Los principales indicadores de desempeño por clase fueron los siguientes:

- **Clase 0 (Fallo reproductivo):** Precisión 0.62, Sensibilidad (recall) 0.54, F1-score 0.58
- **Clase 1 (Entrega exitosa):** Precisión 0.61, Sensibilidad 0.69, F1-score 0.65

A nivel agregado, el modelo mostró un **área bajo la curva ROC (AUC)** de 0.66 (Ilustración 7), lo cual denota una capacidad discriminativa aceptable, aunque inferior a la deseable para aplicaciones clínicas. Este valor sugiere que, ante un par aleatorio de casos (uno exitoso y uno fallido), el modelo tiene un 66% de probabilidad de asignar una mayor probabilidad de éxito al embrión efectivamente entregado.

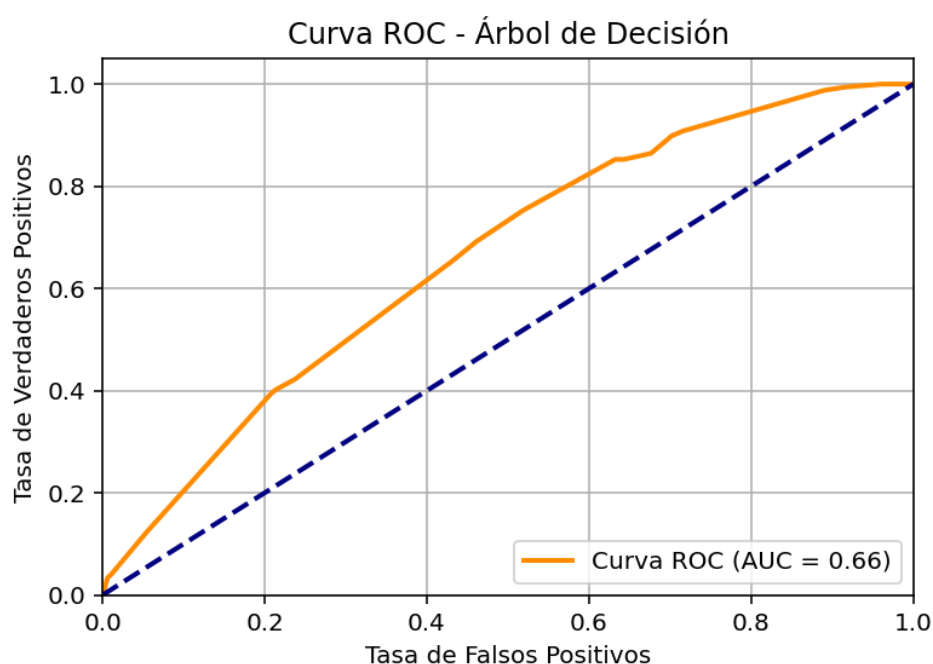


Ilustración 7. Curva ROC del modelo de Árbol de Decisión.

Estos resultados reflejan la capacidad limitada del modelo de árbol para capturar relaciones complejas entre las variables embrionarias y el desenlace reproductivo, y motivan la evaluación de modelos no lineales con mayor capacidad de generalización.

#### 5.1.1.2. Modelo Random Forest sin optimización

Como segundo enfoque, se implementó un modelo de Random Forest con parámetros por defecto ( $n\_estimators = 100$ ,  $max\_depth = 10$ ,  $min\_samples\_split = 5$ ,  $random\_state = 42$ ), a fin de evaluar su desempeño inicial sin ajuste fino. Al igual que en el caso del Árbol de Decisión, se utilizó el mismo conjunto de datos balanceado y preprocesado.

El modelo obtuvo una **precisión global (accuracy)** de 0,637 sobre el conjunto de prueba, mejorando marginalmente el rendimiento del modelo de árbol simple.

La **matriz de confusión** (Tabla 3) muestra un patrón de desempeño más equilibrado entre ambas clases, aunque con cierto sesgo hacia la clase positiva (entrega exitosa). Se observa un aumento en la sensibilidad para la clase 1 ( $recall = 0.77$ ), lo que indica una mayor capacidad para identificar correctamente los embriones que resultarán en potros nacidos.

|      |          | Estimado por el modelo |          |
|------|----------|------------------------|----------|
|      |          | Negativo               | Positivo |
| Real | Negativo | 238                    | 238      |
|      | Positivo | 117                    | 385      |

Tabla 3. Matriz de confusión del modelo Random Forest sin optimización

Los valores específicos de desempeño por clase fueron:

- **Clase 0 (Fallo reproductivo):** Precisión 0.67, Recall 0.50, F1-score 0.57
- **Clase 1 (Entrega exitosa):** Precisión 0.62, Recall 0.77, F1-score 0.68

La evaluación mediante la **curva ROC** (Ilustración 8) arrojó un área bajo la curva (**AUC**) de 0.71, lo cual representa una mejora sustancial respecto del modelo anterior y refleja una mejor

capacidad del modelo para distinguir entre embriones con desenlaces reproductivos exitosos y fallidos.

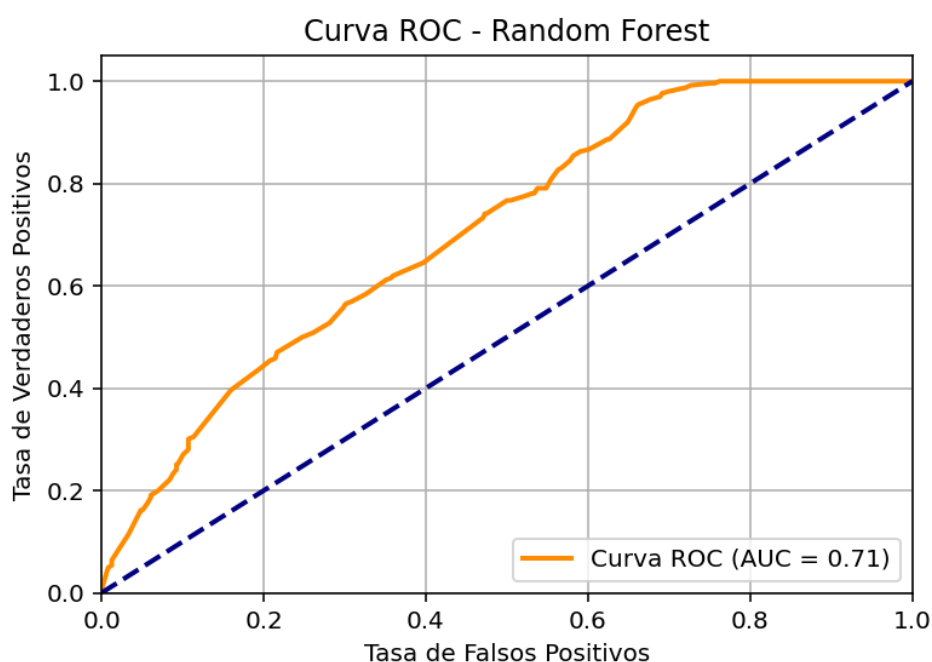


Ilustración 8. Curva ROC del modelo Random Forest sin optimización.

En términos generales, este modelo base de Random Forest demostró ser superior al Árbol de Decisión en todos los indicadores clave, evidenciando la capacidad del ensamblado de árboles para capturar relaciones no lineales y reducir el sobreajuste. No obstante, se espera que una versión optimizada mediante ajuste de hiperparámetros mejore aún más su desempeño.

### 5.1.1.3. Modelo Random Forest optimizado

Con el objetivo de mejorar la capacidad predictiva del modelo Random Forest, se ejecutó una búsqueda aleatoria de 10 iteraciones ( $n\_iter = 10$ ) con validación cruzada estratificada de 3 pliegues ( $cv = 3$ ) y  $random\_state = 42$ . El espacio de búsqueda abarcó  $n\_estimators \in \{100, 200, 300\}$ ,  $max\_depth \in \{5, 10, 15, None\}$ ,  $min\_samples\_split \in \{2, 5, 10\}$ ,  $min\_samples\_leaf \in \{1, 2, 5\}$  y  $bootstrap \in \{True, False\}$ .

A pesar del ajuste, la combinación óptima resultó casi idéntica al modelo base ( $n\_estimators \approx 200$ ,  $max\_depth \approx 10$ ,  $min\_samples\_split = 2$ ,  $min\_samples\_leaf = 1$ ,  $bootstrap = True$ ) y la precisión global se mantuvo en 0,6370, evidenciando una consistencia del desempeño.

La **matriz de confusión** (Tabla 4) fue idéntica al modelo anterior, confirmando la estabilidad del comportamiento del clasificador:

|      |          | Estimado por el modelo |          |
|------|----------|------------------------|----------|
|      |          | Negativo               | Positivo |
| Real | Negativo | 238                    | 238      |
|      | Positivo | 117                    | 385      |

Tabla 4. Matriz de confusión del modelo Random Forest optimizado

Los indicadores por clase permanecieron estables:

- **Clase 0 (Fallo reproductivo):** Precisión 0.67, Recall 0.50, F1-score 0.57

- **Clase 1 (Entrega exitosa):** Precisión 0.62, Recall 0.77, F1-score 0.68

El área bajo la curva ROC también se mantuvo en **0.71**, como se observa en la Ilustración 9, confirmando que no se obtuvo una mejora apreciable en la capacidad discriminativa del modelo.

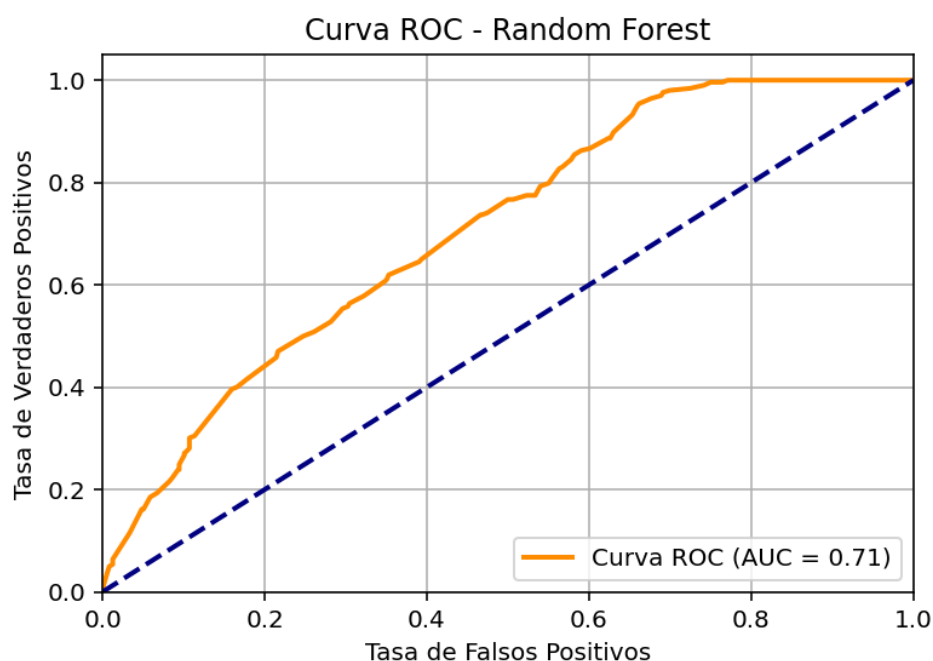


Ilustración 9. Curva ROC del modelo Random Forest optimizado.

En conjunto, estos resultados sugieren que, si bien Random Forest es una técnica eficaz y robusta para este tipo de problemas, la mejora en el desempeño podría requerir el uso de variables adicionales, transformaciones no lineales más sofisticadas o bien modelos con mayor capacidad de ajuste como los basados en boosting.

#### 5.1.1.4. Modelo XGBoost sin optimizar

Como estrategia de aprendizaje de tipo boosting, se implementó el modelo XGBoost utilizando una configuración inicial basada en valores típicos de referencia: 200 árboles ( $n\_estimators=200$ ), profundidad máxima de 10 ( $max\_depth=10$ ), tasa de aprendizaje de 0.1 ( $learning\_rate=0.1$ ) y un factor de ponderación para la clase positiva ( $scale\_pos\_weight=1.5$ ) con el objetivo de atenuar potenciales efectos del desequilibrio de clases. No se aplicó una búsqueda sistemática de hiperparámetros en esta instancia.

El modelo alcanzó una **precisión global (accuracy)** de 0.639, mostrando un rendimiento comparable al de Random Forest, pero con diferencias relevantes a nivel de desempeño por clase.

La **matriz de confusión** (Tabla 5) revela un sesgo marcado hacia la correcta clasificación de casos positivos (embriones que derivaron en entrega exitosa), aunque a costa de una menor sensibilidad para la clase negativa:

|      |          | Estimado por el modelo |          |
|------|----------|------------------------|----------|
|      |          | Negativo               | Positivo |
| Real | Negativo | 197                    | 279      |
|      | Positivo | 74                     | 428      |

Tabla 5. Matriz de confusión del modelo XGBoost sin optimizar

Los indicadores por clase fueron los siguientes:

- **Clase 0 (Fallo reproductivo):** Precisión 0.73, Recall 0.41, F1-score 0.53
- **Clase 1 (Entrega exitosa):** Precisión 0.61, Recall 0.85, F1-score 0.71

El desempeño del modelo en términos de discriminación se visualiza en la **curva ROC**, que arrojó un área bajo la curva (**AUC**) de 0.71 (Ilustración 10). Este valor, equivalente al de los modelos Random Forest, sugiere una adecuada capacidad para diferenciar entre embriones que resultan o no en potros nacidos.

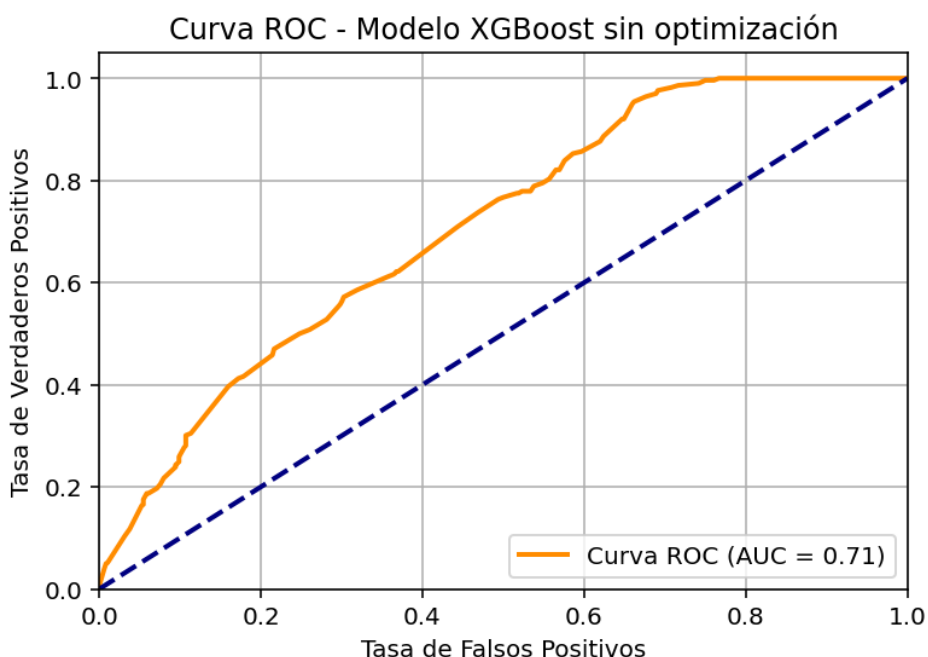


Ilustración 10. Curva ROC del modelo XGBoost sin optimizar.

A pesar de no haber sido ajustado de manera específica, el modelo XGBoost mostró una **mayor sensibilidad para la clase positiva**, lo que se traduce en una mayor capacidad para identificar configuraciones de embriones con alta probabilidad de resultar en un nacimiento exitoso. Esta propiedad lo posiciona como una herramienta valiosa para la estimación probabilística del éxito reproductivo, ya que permite modelar con mayor precisión qué combinaciones de características embrionarias —como tipo, calidad, día de desarrollo y estado de conservación— se asocian con una mayor probabilidad de parto. En este sentido, el objetivo del análisis no es solamente clasificar con exactitud, sino optimizar la asignación de recursos limitados (como las yeguas receptoras) hacia los embriones con mayor potencial de nacimiento.

#### 5.1.1.5. Modelo XGBoost optimizado

Con el objetivo de mejorar la capacidad predictiva del modelo y refinar la estimación de probabilidades de nacimiento, se ajustaron los hiperparámetros de XGBoost mediante validación cruzada manual. El modelo final utilizó 300 árboles ( $n\_estimators=300$ ), una profundidad máxima de 12 ( $max\_depth=12$ ), tasa de aprendizaje reducida ( $learning\_rate=0.03$ ) y un factor de ponderación para la clase positiva elevado ( $scale\_pos\_weight=2.5$ ). Además, se introdujeron restricciones adicionales para evitar sobreajuste: penalización por complejidad ( $gamma=1$ ), y submuestreo tanto de observaciones ( $subsample=0.8$ ) como de predictores por árbol ( $colsample\_bytree=0.8$ ).

El modelo alcanzó una **precisión global (accuracy)** de 0.6483, la más alta de todos los modelos probados. No obstante, su principal fortaleza radica en su **extraordinaria sensibilidad para predecir nacimientos**, con una tasa de recall del 98% para la clase positiva.

La **matriz de confusión** (Tabla 6) revela una estrategia de clasificación altamente conservadora ante falsos negativos:

|      |          | Estimado por el modelo |          |
|------|----------|------------------------|----------|
|      |          | Negativo               | Positivo |
| Real | Negativo | 144                    | 332      |
|      | Positivo | 12                     | 490      |

Tabla 6. Matriz de confusión del modelo XGBoost optimizado

Esto se traduce en los siguientes indicadores por clase:

- **Clase 0 (Fallos reproductivos):** Precisión 0.92, Recall 0.30, F1-score 0.45
- **Clase 1 (Nacimientos):** Precisión 0.60, Recall 0.98, F1-score 0.74

Si bien el modelo incurre en un mayor número de falsos positivos, su altísima capacidad para detectar correctamente los embriones con potencial de nacimiento lo convierte en una herramienta prioritaria para la toma de decisiones reproductivas. Este enfoque es coherente con el objetivo del estudio, centrado en **maximizar la probabilidad de nacimiento**, aún a costa de admitir errores tipo I (clasificar como viable un embrión que no lo es), siempre que se minimicen los errores tipo II (descartar erróneamente un embrión que sí habría nacido).

En este sentido, la **ligera reducción del área bajo la curva ROC (AUC = 0.70)** que se muestra en la Ilustración 11 en comparación con otros modelos (que alcanzaron hasta 0.71) **no debe interpretarse como una pérdida significativa de desempeño**, sino como un ajuste en favor de la sensibilidad. Dado que el AUC refleja un equilibrio global entre sensibilidad y especificidad, este pequeño descenso es esperable —y aceptable— cuando se prioriza la correcta detección de nacimientos por sobre la clasificación estricta de ambas clases.

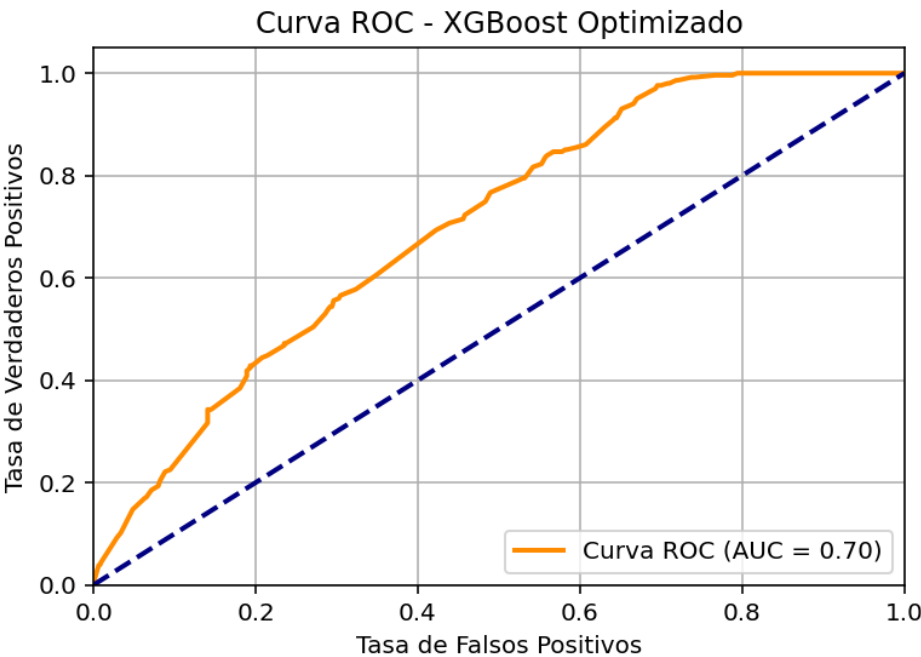


Ilustración 11. Curva ROC del modelo XGBoost optimizado.

En conjunto, el modelo XGBoost optimizado se configura como la mejor herramienta predictiva del estudio, permitiendo **estimar con alto nivel de confianza qué combinaciones de tipo de embrión, calidad, día y estado de conservación conducen a un nacimiento exitoso**. Esta capacidad analítica representa un aporte valioso tanto para la mejora de decisiones clínicas como para la asignación estratégica de yeguas receptoras, orientada a la eficiencia reproductiva.

### 5.1.2. Combinaciones Óptimas para Maximizar la Probabilidad de Nacimiento

El análisis del modelo XGBoost optimizado permitió identificar las configuraciones de variables embrionarias con mayor probabilidad estimada de resultar en nacimiento. Estas combinaciones, presentadas a continuación en la Tabla 7, ofrecen una valiosa guía para la toma de decisiones cuando existen múltiples embriones disponibles para transferencia:

| Ranking | Tipo de embrión   | Calidad | Conservación | Día | Probabilidad estimada |
|---------|-------------------|---------|--------------|-----|-----------------------|
| 1       | Mórula            | G2.5    | Vitrificado  | 8.5 | 0.9407                |
| 2       | Mórula            | G2.5    | Vitrificado  | 8.0 | 0.9407                |
| 3       | Blastocisto Prot. | G2      | Vitrificado  | 7.5 | 0.9034                |

*Tabla 7. Combinaciones Óptimas para Maximizar la Probabilidad de Nacimiento*

Estas probabilidades deben ser interpretadas con criterio. **No constituyen garantías individuales de éxito**, sino que representan **estimaciones estadísticas derivadas de patrones observados en embriones con características similares** dentro del conjunto de datos utilizado para entrenar el modelo. Por lo tanto, una probabilidad estimada del 94% no implica que un embrión con dicha configuración nacerá con certeza. Lo que sugiere es que, **en un escenario en el que se transfirieran múltiples embriones con esas mismas características y bajo condiciones similares, se esperaría que aproximadamente el 94% resultaran en nacimiento**.

Sin embargo, la concreción o no de un nacimiento en un caso puntual sigue sujeta a incertidumbre, ya que el modelo no contempla todas las variables que influyen en el desenlace reproductivo. En particular, uno de los factores más relevantes no modelados es **la genética de la línea clonada**, la cual, según observaciones internas y registros históricos del laboratorio, explica aproximadamente **un 35% a 40% de la variabilidad en los resultados de nacimiento**. Este componente genético no está representado entre las variables utilizadas en el modelo, lo que limita parcialmente su capacidad predictiva y explica, al menos en parte, su desempeño general (accuracy = 64,7%).

**Esta omisión introduce una posible fuente de sesgo en la interpretación de los resultados**. Es decir, algunas de las combinaciones óptimas identificadas por el modelo podrían coincidir, por azar o correlación estructural, con líneas genéticas que presentan una mayor predisposición natural al nacimiento. En tal caso, el modelo podría estar atribuyendo el éxito a las características visibles del embrión (como calidad o día de evolución), cuando en realidad estaría captando de forma indirecta el efecto de la genética subyacente. Esta limitación no invalida la utilidad del modelo, pero sí exige **cautela adicional al interpretar sus predicciones como causalmente explicativas**.

Adicionalmente, debe considerarse que el laboratorio no siempre dispone de embriones con estas configuraciones ideales. Las características observadas (tipo, calidad, día y estado de conservación) **no son completamente controlables**, ya que dependen del desarrollo natural del embrión y de las condiciones particulares del proceso de clonación. Por lo tanto, si bien estas configuraciones representan un ideal teórico, **su disponibilidad no puede garantizarse**, y su aparición es eventual.

En consecuencia, la utilidad práctica de esta información reside en la **priorización relativa**: ante la existencia de múltiples embriones disponibles para transferencia, aquellos que se aproximen a estas combinaciones óptimas podrán ser priorizados, sin que ello implique una valoración económica diferenciada ni una modificación operativa en la producción. Este enfoque **maximiza la eficiencia del sistema sin alterar su lógica productiva**, y sin generar expectativas infundadas sobre el rendimiento individual de cada embrión.

### 5.1.3. Implicancias para la gestión estratégica

Desde una perspectiva de gestión reproductiva, el modelo desarrollado brinda herramientas prácticas para **mejorar la asignación de recursos limitados, como las yeguas receptoras**, y para introducir criterios objetivos en procesos de alta variabilidad.

En particular, permite:

- **Priorizar embriones con mayor probabilidad estimada de nacimiento**, mejorando la eficiencia global del sistema sin aumentar los costos ni alterar la operación diaria.
- **Sistematizar la toma de decisiones** ante situaciones de sobreoferta de embriones o cupo limitado de receptoras, evitando decisiones subjetivas o arbitrarias.
- **Documentar y justificar decisiones clínicas** de forma replicable y auditada, lo que aporta transparencia y trazabilidad al proceso.

Al mismo tiempo, el modelo refuerza un cambio de paradigma: **pasar de decisiones binarias basadas en criterios fijos (ej., "día 7 es mejor que día 8") hacia decisiones probabilísticas personalizadas**, adaptadas a la configuración integral de cada embrión.

## 5.2. Simulación de escenarios de reducción del número de transferencias

El análisis predictivo de eficiencia individual no es suficiente para definir políticas operativas, ya que la asignación de embriones ocurre dentro de un sistema con restricciones. Por este motivo, se desarrolló una simulación estocástica con el objetivo de complementar el estudio individual y evaluar **cómo varía el rendimiento global del sistema ante diferentes cantidades máximas de transferencias por clon solicitado**, en términos de entregas, eficiencia y retorno económico.

### 5.2.1. Lógica de la simulación

Partiendo de la base de que se asignan múltiples embriones por clon para asegurar al menos una entrega, se construyó una simulación Montecarlo que incorpora:

- **Probabilidad empírica de éxito por línea genética**, derivada de los datos históricos.
- Capacidad fija de 1.000 receptoras por simulación.
- Reglas de ingresos escalonados según cantidad de nacimientos por clon.
- Costos variables asociados a transferencias y producción de embriones.

La simulación explora escenarios desde 1 hasta 10 transferencias por clon, buscando responder dos preguntas clave:

1. ¿Existe un punto óptimo de eficiencia donde se maximizan los nacimientos sin incurrir en sobrecostos?

## 2. ¿Cómo varían los indicadores económicos, comerciales y operativos clave según el protocolo aplicado?

Este enfoque permite ir más allá del análisis basado únicamente en **promedios históricos**, incorporando la **variabilidad real observada en la eficiencia de las distintas líneas genéticas**. Además, permite evaluar el impacto de diferentes políticas de asignación de transferencias sobre los resultados reproductivos y económicos, bajo condiciones operativas realistas tales como la **capacidad fija de receptoras y los costos asociados**.

### 5.2.2. Clones intentados y exitosos

Un primer indicador clave es la relación entre clones intentados (según la capacidad fija de receptoras) y clones exitosos (nacimientos con al menos una entrega) llamada eficiencia. La Ilustración 12 muestra que a medida que se reduce el número de transferencias por clon, aumenta la cantidad de líneas que pueden ser intentadas (mayor cobertura genética), pero también el riesgo de que caer en bajas eficiencias.

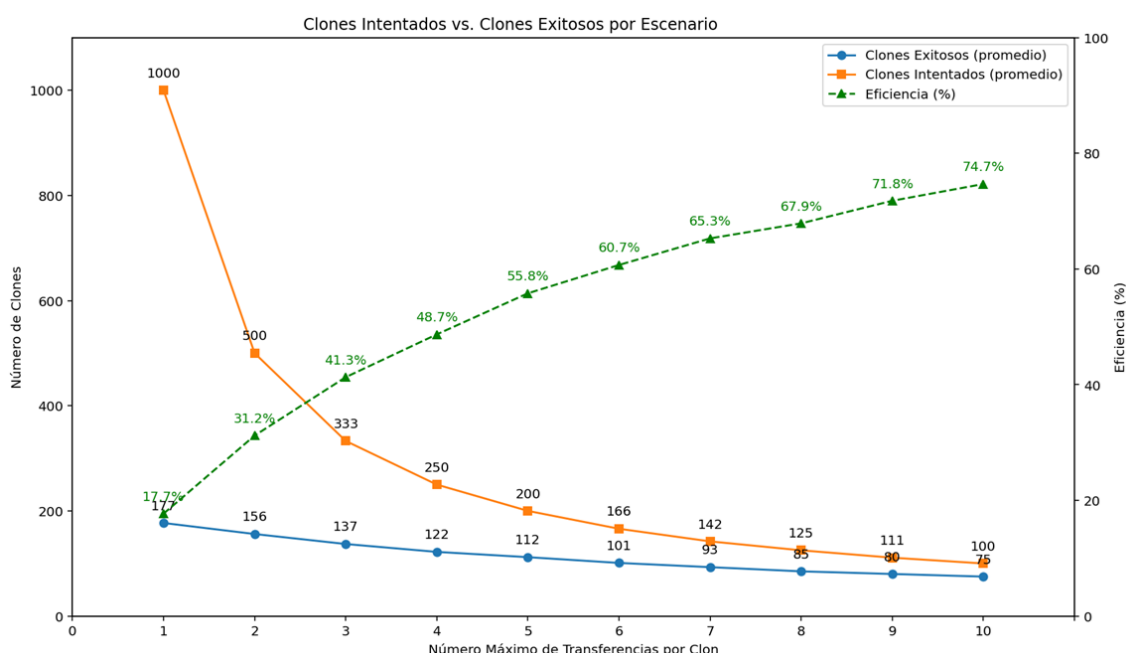


Ilustración 12. Relación entre clones intentados y clones exitosos por escenario (1 a 10 transferencias).

En el extremo de una sola transferencia por clon pedido, el protocolo maximiza la cobertura (1.000 líneas por simulación), pero compromete la eficiencia y, por ende, el cumplimiento de los contratos. En cambio, con 7 u 8 transferencias por línea, la entrega resulta mucho más confiable, aunque solo se cubren unas 150 líneas por simulación. Estos resultados evidencian la tensión estructural entre eficiencia operativa, cobertura genética y gestión comercial.

### 5.2.3. Costo por clon exitoso

Con el objetivo de preservar la confidencialidad de la estructura de costos, se presenta el análisis mediante un **índice relativo base 100**. Este índice se construye tomando como referencia el costo unitario por clon exitoso del primer escenario evaluado (menor cantidad de transferencias por clon), al cual se le asigna un valor de 100. Cada escenario posterior refleja el costo relativo en porcentaje respecto a esa base inicial.

La fórmula utilizada es (Ec. 1):

$$\text{Índice de Costo Relativo} = \left( \frac{\text{Costo Unitario por Clon Exitoso}}{\text{Costo Inicial}} \right) * 100 (\text{Ec. 1})$$

El análisis económico comienza con el **costo unitario por clon exitoso**, resultado de dividir el costo total por la cantidad de entregas. La Ilustración 13 muestra una curva con mínimo local en 3 transferencias por clon, donde se logra el mejor balance entre intentos y costos:

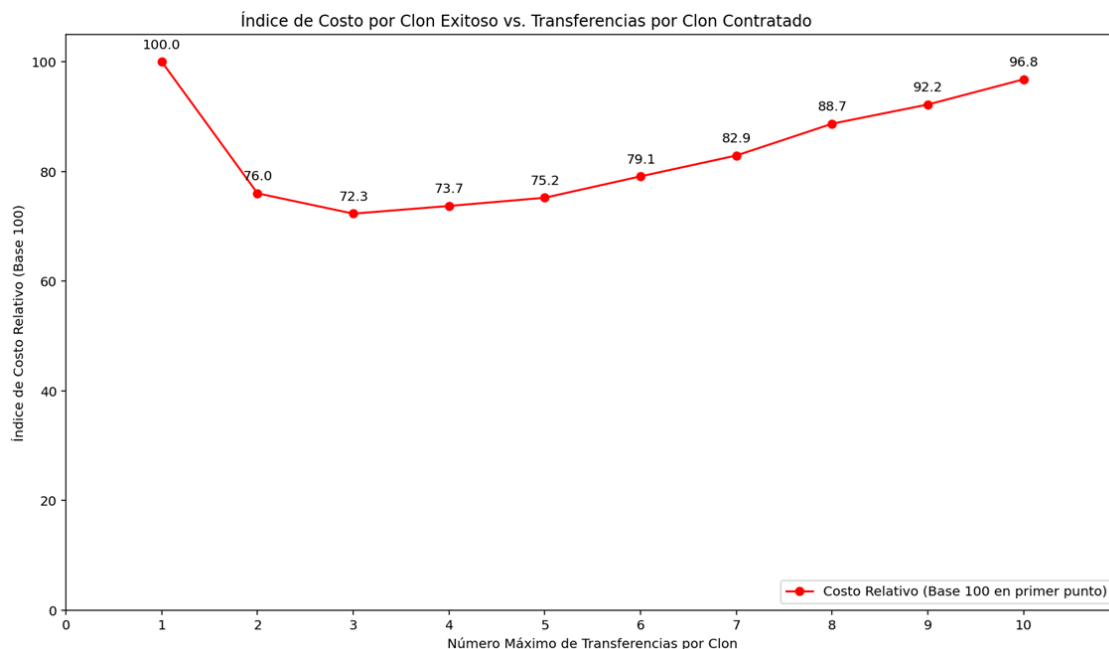


Ilustración 13. Índice de Costo por Clon Exitoso vs. Transferencias por Clon Contratado

Este resultado indica que el rango entre 2 y 5 clones genera costos unitarios relativamente controlados. Por fuera de este intervalo, los costos unitarios aumentan y podrían no ser eficientes para el negocio.

#### 5.2.4. Rentabilidad neta y retorno global

La Ilustración 14 presenta la evolución conjunta del ingreso total, costo total y rentabilidad neta relativa promedio para cada escenario:

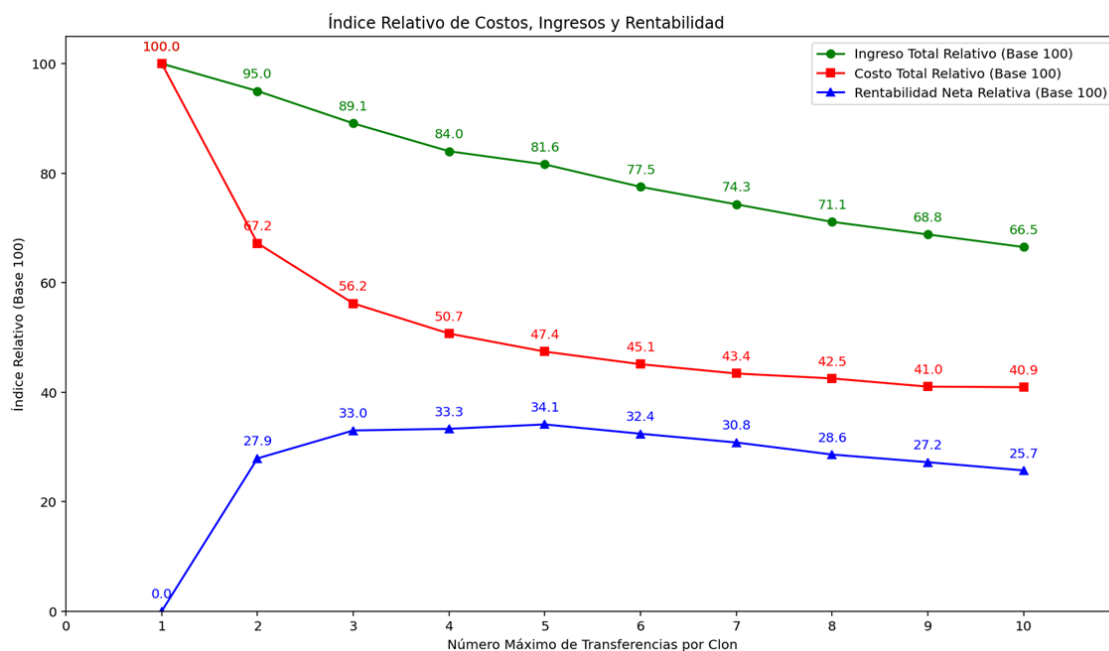


Ilustración 14. Curva de ingresos, costos y rentabilidad neta relativa según cantidad de transferencias por clon

La rentabilidad neta alcanza su valor máximo cuando se asignan **cinco transferencias por clon, superando en un 10,7% la rentabilidad obtenida con siete transferencias**. Este escenario combina buenos niveles de entrega con un número moderado de intentos y un control más eficiente de los costos, respaldando así la hipótesis de que una estrategia de diversificación moderada resulta más rentable que una estrategia de sobreasignación. Además, esta mejora se traduce en un **aumento del margen neto de 33,56 puntos básicos (bps)**.

#### 5.2.5. Eficiencia integral: Commercial Benefit Ratio (CBR)

Para integrar las dimensiones económicas y comerciales (consecuencia de la dimensión operativa), se definió un índice de eficiencia adimensional, el **Commercial Benefit Ratio (CBR)**, que pondera la rentabilidad neta por la tasa de éxito del escenario (Ec.2). Este índice fue reconvertido a base 100 como todos los índices para resguardar la confidencialidad de la estructura económica de la empresa.

$$\text{Commercial Benefit Ratio} = \% \text{Clones exitos} * \text{Rentabilidad neta}(\text{Ec. 2})$$

El máximo de esta función otorga una combinación óptima entre satisfacción del cliente y rentabilidad del negocio.

La Ilustración 15 muestra que el CBR se maximiza con **cinco transferencias**, y decrece a partir de seis o más:

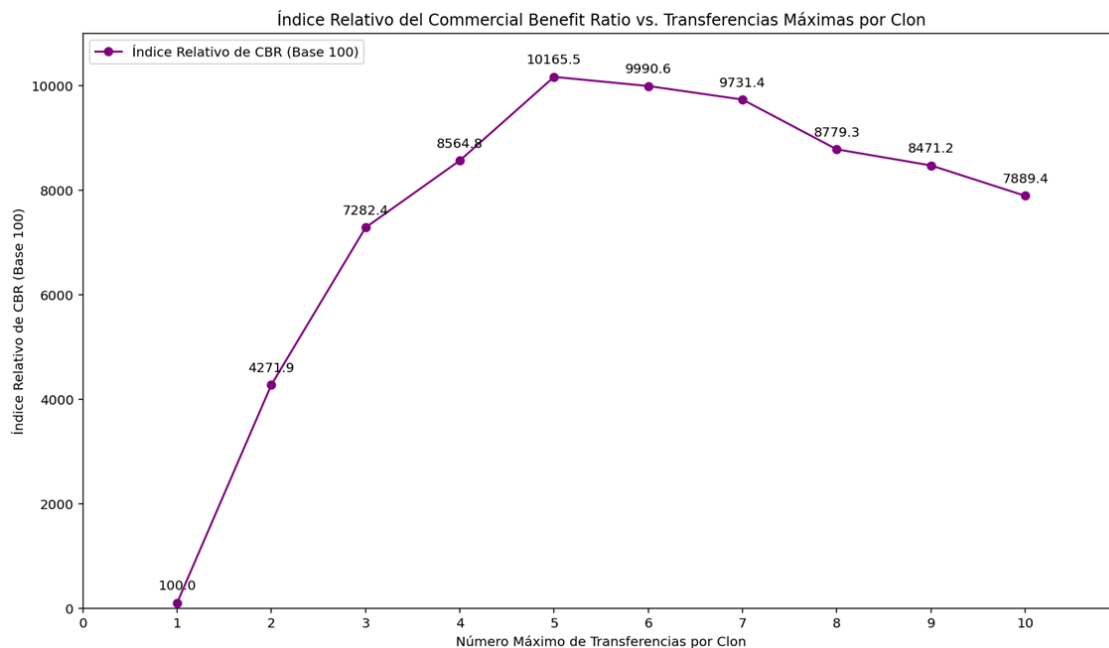


Ilustración 15. Índice Relativo del Commercial Benefit Ratio vs. Transferencias Máximas por Clon

Este indicador confirma que el esquema actual (7 transferencias) **no es óptimo**: el sistema alcanza su mejor rendimiento integral antes de ese punto, exactamente con 5. A partir de allí, los ingresos adicionales por más nacimientos no compensan los costos extra ni el uso ineficiente de receptoras.

#### 5.2.6. Implicancias prácticas

Los resultados de la simulación permiten **cuestionar la política actual de asignar siete transferencias por clon como estándar fijo**, ya que no surge como el valor más eficiente bajo ninguna dimensión evaluada. Tanto el análisis de costos unitarios, como la rentabilidad neta y el índice CBR, convergen en señalar que el número óptimo se encuentra en torno a **cinco transferencias por clon**.

Dado que en la práctica **las líneas celulares rara vez se repiten entre clones solicitados, no es viable establecer un protocolo diferenciado por línea genética**. Esto implica que, lejos de aplicar una regla adaptativa según eficiencia histórica, el enfoque más racional es **reducir el número de transferencias por clon de manera uniforme**, con el objetivo de **maximizar la diversificación del riesgo genético**.

Reducir la asignación de embriones a cinco por clon permitiría:

- **Incrementar la cantidad de líneas diferentes cubiertas** con el mismo stock de embriones y receptoras.
- **Reducir el costo por clon exitoso**, que alcanza su punto más bajo en las cercanías de este valor.
- **Maximizar la rentabilidad neta y el índice CBR**, sin afectar significativamente la tasa de entregas.

Esta estrategia no sólo es más eficiente en términos económicos y operativos, sino que también responde a un objetivo clave del centro: **aumentar la cobertura genética efectiva**, reduciendo la dependencia de un promedio histórico general y adoptando un enfoque basado en evidencia.

### 5.3. Evaluación económica y estratégica

La evidencia empírica generada en este trabajo permite revisar críticamente la lógica operativa del protocolo vigente en el centro de clonación, que asigna **siete transferencias por clon solicitado como política uniforme**. Tanto los modelos predictivos como los escenarios simulados sugieren que dicha regla, aunque conservadora, no representa el punto de máxima eficiencia operativa ni económica.

#### 5.3.1. Síntesis de hallazgos

El análisis predictivo reveló que ciertas combinaciones de tipo, calidad, día y estado de conservación del embrión presentan **probabilidades de éxito superiores al azar**. Sin embargo, como estas características no son configurables a voluntad, su utilidad está en orientar decisiones bajo disponibilidad abundante, no en construir reglas fijas.

Así, más decisiva fue la simulación de escenarios operativos con diferentes números máximos de transferencias por clon.

#### 5.3.2. Restricciones reales y criterios de decisión

En el centro de clonación **la capacidad reproductiva es limitada por la cantidad de yeguas receptoras**, y **la diversidad genética dada por una amplia cartera de clientes es un activo estratégico**. Además, como las líneas celulares no se repiten con frecuencia, es decir, no se vuelven a clonar los mismos donantes, **no puede aplicarse una política diferenciada por línea basada en eficiencia histórica**.

Por lo tanto, el número de transferencias por clon no debe pensarse únicamente en función de maximizar la tasa de éxito, sino también de **maximizar la cobertura genética con los recursos disponibles**, diversificando el riesgo de falla entre líneas independientes.

Desde esta perspectiva, reducir el número de transferencias a cinco implica:

- Aumentar la cantidad de líneas que pueden ser cubiertas con el mismo número de receptoras.
- Mantener una tasa de entregas aceptable, cercana a la observada con siete transferencias.
- Minimizar los excedentes, reduciendo el número de clones no deseados que generan costos adicionales o presiones comerciales.

#### 5.3.3. Implicancias estratégicas

Más allá de su impacto operativo, este cambio de protocolo tiene implicancias estratégicas relevantes para el modelo de negocio del centro de clonación:

- **Fortalece la sostenibilidad económica** del sistema, al reducir costos unitarios sin comprometer ingresos esperados y al aumentar los márgenes.
- **Alinea la toma de decisiones con evidencia empírica**, reemplazando el criterio histórico por un análisis basado en eficiencia comparada.

Esta revisión operativa no implica una ruptura con el enfoque técnico del laboratorio, sino una **evolución en la política de asignación**, coherente con los nuevos datos disponibles y con las metas comerciales del negocio.

## 6. Discusión

### 6.1. Redefinición de la lógica operativa en contextos de alta incertidumbre

La práctica de la clonación equina opera sobre una base biológica compleja, con múltiples fuentes de variabilidad no controlables. Tradicionalmente, este entorno de alta incertidumbre se ha abordado mediante protocolos conservadores, como la asignación de un número fijo de transferencias por clon, basados en promedios históricos de éxito.

Sin embargo, los resultados de esta tesis permiten proponer una **redefinición del criterio operativo**, introduciendo un enfoque más granular y eficiente, basado en evidencia cuantitativa. Lejos de aplicar reglas homogéneas, el modelo sugiere que es posible —y deseable— ajustar las decisiones reproductivas a partir de estimaciones empíricas de eficiencia marginal y simulaciones de escenarios reales.

Este cambio no implica asumir riesgos innecesarios, sino **administrar mejor los recursos escasos**, como las yeguas receptoras, y ampliar el alcance estratégico del centro, priorizando la cobertura de líneas genéticas diversas con base en criterios objetivos.

### 6.2. Aportes del enfoque analítico al negocio de la clonación equina

Uno de los principales aportes de este trabajo radica en haber demostrado que herramientas de análisis predictivo y simulación pueden integrarse eficazmente a un entorno técnico de alta especialización. Hasta ahora, las decisiones reproductivas en la clonación han estado mayormente guiadas por el juicio experto y el conocimiento acumulado por los profesionales del laboratorio. Si bien esta experiencia es valiosa e insustituible, su combinación con modelos formales puede mejorar significativamente la toma de decisiones. Además, dado que el campo de la clonación equina se encuentra aún en una etapa incipiente de investigación y desarrollo, sin haber alcanzado una madurez tecnológica plena, existe un espacio relevante para la aplicación de técnicas analíticas que permitan optimizar procesos y potenciar los resultados.

En particular, esta tesis introduce y valida un modelo que no sólo predice la probabilidad de éxito embrionario con un rendimiento aceptable, sino que además permite simular escenarios estratégicos con distintas asignaciones de recursos, generando métricas clave como el costo por clon exitoso, la rentabilidad neta y un índice sintético como el Commercial Benefit Ratio (CBR).

Este tipo de herramientas, cuando son correctamente interpretadas, **no reemplazan la experiencia del laboratorio, pero sí permiten sistematizarla, contrastarla con evidencia histórica y actualizarla ante nuevas condiciones**. En este sentido, abren un nuevo campo dentro de la gestión reproductiva basada en datos, con aplicaciones no sólo en clonación equina, sino potencialmente en otras biotecnologías reproductivas o en centros que operen con restricciones similares.

### 6.3. Limitaciones del estudio

Como todo trabajo empírico, esta investigación presenta limitaciones que deben ser consideradas al interpretar sus resultados:

- **Variables no observadas:** el modelo predictivo no incorpora información sobre la receptora (estado hormonal, antecedentes reproductivos, salud), ni sobre las condiciones específicas del entorno de transferencia. Estas omisiones pueden explicar parte de la variabilidad que el modelo no logra capturar, y representan oportunidades claras de mejora para futuras versiones del modelo.
- **Estructura retrospectiva de los datos:** el análisis se basó en registros históricos, lo que implica que los datos ya contienen las decisiones tomadas por el laboratorio en el pasado. Esto introduce posibles sesgos, como una sobrerrepresentación de ciertas líneas genéticas o tipos de embriones que hayan sido preferidos operativamente.
- **Evaluación simulada de escenarios:** si bien la simulación Montecarlo fue rigurosamente diseñada, los escenarios aún no han sido implementados en la práctica. La validación empírica de la política sugerida (cinco transferencias por clon) en una temporada completa permitiría confirmar la efectividad del modelo en condiciones reales. Esto será implementado a partir de la próxima temporada (Jul 25-Jun 26)

#### 6.4. Lineamientos para futuras investigaciones

Los resultados abren múltiples líneas de trabajo futuras. Entre ellas, se destaca como especialmente prometedora la posibilidad de **explorar la relación entre eficiencia reproductiva y genealogía genética en la raza Polo Argentino**. Existen líneas fundadoras reconocidas en la genealogía de estos caballos, y su influencia podría correlacionarse con diferencias sistemáticas en los resultados reproductivos.

Integrar **árboles genealógicos profundos** al análisis permitiría:

- Identificar si ciertos linajes fundadores están asociados a tasas de éxito significativamente más altas o más bajas.
- Detectar subgrupos genéticos con eficiencia reproductiva diferenciada, lo cual sería de alto valor comercial.
- Complementar el enfoque predictivo con un análisis estructural de herencia, utilizando herramientas de redes genealógicas o clústeres filogenéticos.

Además, futuras investigaciones podrían incluir:

- Datos clínicos de las receptoras y factores ambientales del entorno de transferencia.
- Desarrollo de herramientas operativas que combinen el modelo predictivo con métricas de eficiencia para uso en tiempo real.
- *Benchmarking* con otros centros reproductivos para validar la generalización del enfoque.

#### 6.5. Reflexión final

Esta tesis propone una **nueva forma de pensar la gestión reproductiva en clonación equina de este centro**, combinando datos históricos, modelos de aprendizaje automático y simulación estratégica. En lugar de adherir pasivamente a un protocolo establecido por promedios, el trabajo sugiere que **es posible tomar decisiones más eficientes, más justas y más alineadas con los objetivos del negocio**, sin perder rigor técnico ni operatividad.

En contextos donde los recursos son escasos y la incertidumbre es elevada, **no es posible controlar todos los factores, pero sí es posible decidir mejor**. Y eso —en última instancia— es lo que distingue a un sistema inteligente: su capacidad de aprender, adaptarse y evolucionar.

## 7. Conclusión

### 7.1. Síntesis de objetivos y hallazgos

El propósito central de esta tesis fue **evaluar la eficiencia del protocolo de asignación de embriones en un centro de clonación equina** que, hasta el momento, operaba bajo un esquema homogéneo de siete transferencias por clon solicitado. A través de un enfoque mixto que combinó **análisis exploratorio de datos, modelos predictivos de machine learning y simulación estocástica**, se buscó responder si era posible diseñar una política alternativa más eficiente en términos reproductivos, económicos y estratégicos.

El trabajo comenzó con el análisis exploratorio de los datos para poder comprender qué características podían estudiarse del proceso. Posteriormente, se procedió con la construcción de un modelo predictivo que, a partir de variables observables del embrión (tipo, calidad, día de evolución, estado de conservación), estima la probabilidad de que dicha transferencia resulte en una entrega. El modelo XGBoost mejorado alcanzó un **recall del 98 % para los casos exitosos** y un accuracy general del 65 %, valores consistentes con el nivel de incertidumbre biológica reconocido por el laboratorio. La importancia relativa de las variables reveló que ciertos perfiles embrionarios presentan probabilidades de éxito superiores al 90 %, configurando un conjunto de combinaciones óptimas.

Si bien no es posible manipular las variables para obtener, por ejemplo, un embrión G2.5 en día 8.5, en algunos casos sí puede seleccionarse, entre dos embriones disponibles, aquel con mejores características para transferir. Dado que esta capacidad de elección es limitada y no modifica la disponibilidad general de embriones, se desarrolló una segunda etapa de análisis basada en simulaciones de escenarios operativos, considerando diferentes cantidades máximas de transferencias por clon. El análisis incorporó restricciones reales, como la capacidad fija de yeguas receptoras y los costos asociados a la producción y transferencia de embriones. El resultado fue contundente: el número de cinco transferencias por clon maximizó el costo-beneficio global del sistema, alcanzando el mejor resultado en términos de:

- Rentabilidad neta promedio: un 10,7% más que en el escenario actual.
- Costo por clon exitoso: si bien no es el óptimo en términos de las simulaciones, se lograría reducir en un 9,3% respecto a las 7 transferencias.
- Commercial Benefit Ratio (CBR): un 4,5% mayor al actual.

Esto sugiere que el protocolo actual no es eficiente y que su revisión podría traer beneficios sustanciales, sin afectar significativamente la tasa global de entregas.

### 7.2. Contribuciones prácticas y metodológicas

Desde el punto de vista práctico, esta tesis ofrece al centro de clonación una herramienta concreta para **repensar sus decisiones reproductivas**, que hasta ahora se basaban en criterios promedio no diferenciados. El trabajo demuestra que, aún en contextos biológicos con incertidumbre irreducible, **es posible introducir eficiencia mediante el análisis de datos históricos combinados con simulación y modelos de predicción**.

En términos metodológicos, el enfoque busca innovar en el rubro: se combinó aprendizaje supervisado, simulación Montecarlo y métricas económico-operativas en un único marco analítico integrado. Este diseño permite **pasar de la predicción individual al planeamiento estratégico**, respetando tanto las limitaciones biológicas como las restricciones de capacidad del sistema.

Además, la construcción de indicadores sintéticos como el CBR permite traducir conceptos complejos (como la rentabilidad ajustada por éxito reproductivo) en métricas accionables, facilitando su comunicación a equipos no técnicos.

### 7.3. Implicancias estratégicas y de gestión

La revisión del número de transferencias por clon no es solo una cuestión técnica, sino también una **decisión de política operativa** con impactos estratégicos. Implementar un protocolo basado en cinco transferencias permitiría:

- Diversificar el riesgo genético del proceso,
- Reducir costos operativos,
- Disminuir la probabilidad de excedentes no deseados (clones que deben ser descartados o comercializados sin planificación),
- Mejorar la eficiencia percibida del servicio por parte de los clientes.

Desde una perspectiva más amplia, este tipo de decisiones posiciona al centro como una organización **orientada a datos, adaptativa y con foco en la rentabilidad sustentable**, cualidades cada vez más valoradas en entornos competitivos de alta tecnología.

### 7.4. Límites del trabajo y perspectivas futuras

Si bien los hallazgos son consistentes y robustos, esta tesis reconoce ciertas **limitaciones**:

- El modelo predictivo no incorpora variables críticas del lado de la receptora, como sincronización hormonal, edad, o antecedentes reproductivos.
- La estructura de los datos es retrospectiva y puede contener sesgos no controlados.
- La implementación del protocolo propuesto aún no ha sido testeada en una temporada real.

A futuro, sería valioso **ampliar la base de datos con información de receptoras**, así como desarrollar una **temporada operativa con cinco transferencias por clon**, que permita comparar los resultados proyectados con los reales.

Además, se identificó como línea prioritaria de investigación el estudio de **eficiencia reproductiva en relación con los linajes fundadores del Polo Argentino**. El análisis de árboles genealógicos profundos podría revelar estructuras latentes en los datos que ayuden a predecir comportamientos reproductivos más allá de las características fenotípicas del embrión.

### 7.5. Reflexión final

Esta tesis buscó mejorar la eficiencia de un proceso biotecnológico clave, no solo desde una perspectiva técnica, sino también estratégica y económica. Demostró que, aún en contextos

de alta incertidumbre como la clonación equina, es posible tomar mejores decisiones cuando se combina experiencia con datos y modelos rigurosos.

El cambio propuesto puede parecer menor en magnitud, pero representa un giro conceptual significativo: pasar de un criterio homogéneo y empírico a una lógica de asignación basada en eficiencia marginal y riesgo diversificado.

En definitiva, el aporte de este trabajo no es solo una mejora operativa. Es la invitación a repensar cómo se decide, cómo se usa la información disponible y cómo se puede construir valor desde la inteligencia aplicada. Un camino que, sin duda, recién empieza.

Desde una perspectiva personal, la formación en el Master in Management & Analytics de la Universidad Torcuato Di Tella resultó decisiva para abordar este desafío con una mirada diferente. Este programa combina la rigurosidad analítica con una fuerte orientación a la toma de decisiones estratégicas, brindando herramientas para transformar datos en acciones de alto impacto. Aplicar ese enfoque a un problema de biotecnología reproductiva no solo amplió mis competencias técnicas, sino que reafirmó el valor de integrar ciencia de datos y gestión estratégica en sectores donde tradicionalmente predominaba la experiencia empírica.

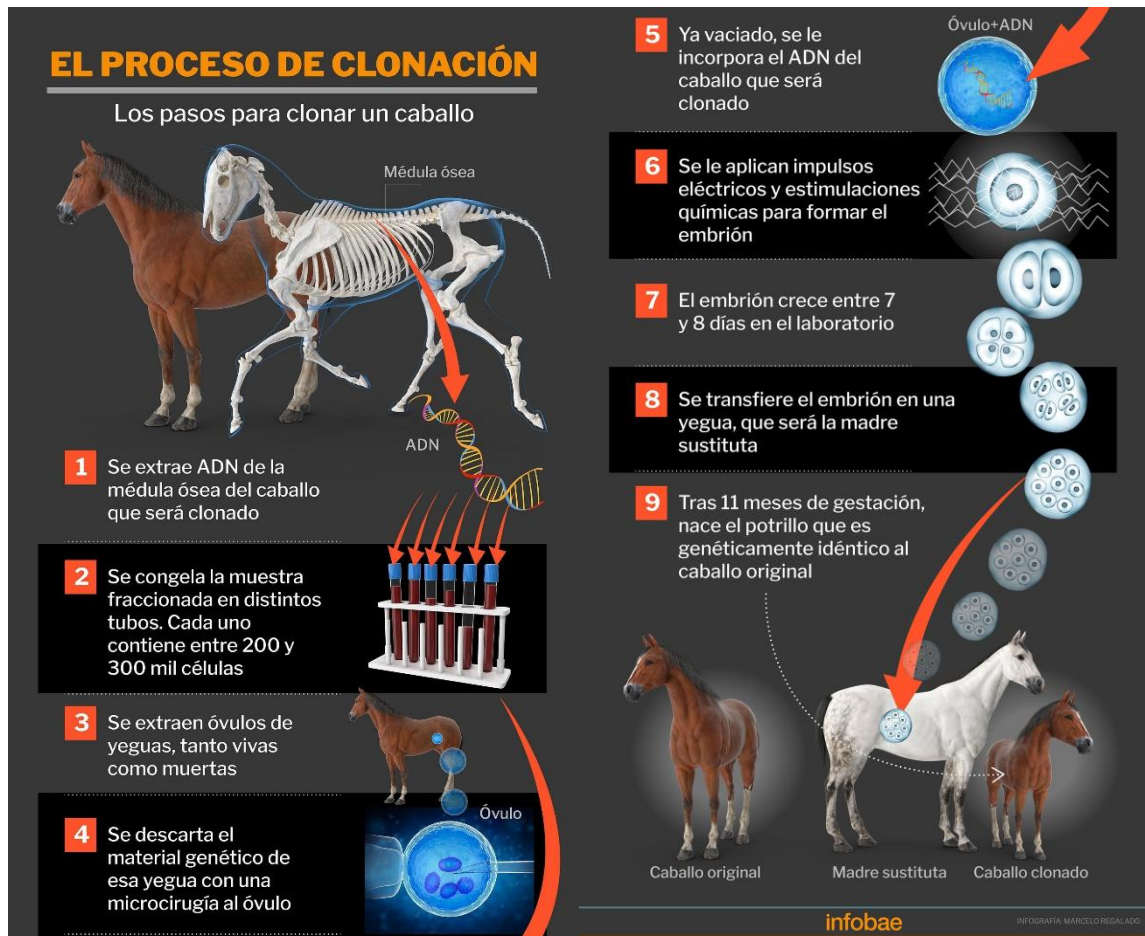
Esta tesis es, en parte, una síntesis de ese recorrido: un ejercicio de pensamiento crítico, modelización cuantitativa y visión aplicada que aspira a generar un impacto real en la práctica profesional.

## Referencias

1. Carrillo Sánchez, S. A. (2021). *Análisis de las variables que afectan la eficiencia de un programa de transferencia embrionaria equina en una producción comercial de embriones en la Sabana de Bogotá*. Universidad de los Llanos.
2. Galli, C. (2013). ARTs in large animals: Achievements and unmet promises. *Animal Reproduction*, 14(3), 614–621.
3. Galli, C., Colleoni, S., Duchi, R., Lagutina, I., & Lazzari, G. (2013). Advanced assisted reproduction in the horse. *Animal Reproduction*, 10(3), 334–343.
4. Grady, S. T., & Hinrichs, K. (2016). Factores que afectan la eficiencia de la producción de potros por inyección intracitoplasmática de espermatozoides (ICSI). *Spermova*, 6(1), 18–21. <https://doi.org/10.18548/aspe/0003.03>
5. Oliveira, R. A., Alonso, M. A., Fonte, J. S., & Fernandes, C. B. (2024). Equine ICSI: An update on semen perspective. *Animal Reproduction*, 21(4), e20240015. <https://doi.org/10.1590/1984-3143-AR2024-0015>
6. Suvá, M., Arnold, V. H., & Wiedenmann, E. A. (2023). First sex modification case in equine cloning. *PLoS ONE*, 18(1), e0279869. <https://doi.org/10.1371/journal.pone.0279869>
7. Umair, M., Scheeren, V. F. d. C., Beitsma, M. M., Colleoni, S., Galli, C., Lazzari, G., de Ruijter-Villani, M., Stout, T. A. E., & Claes, A. (2023). In vitro-produced equine blastocysts exhibit greater dispersal and intermingling of inner cell mass cells than in vivo embryos. *International Journal of Molecular Sciences*, 24(11), 9619. <https://doi.org/10.3390/ijms24119619>
8. Vélez, I. C., & Hinrichs, K. (2011). Técnicas de reproducción asistida en caballos. *Revista CES Medicina Veterinaria y Zootecnia*, 6(2), 124–128.
9. The Washington Post. (2023). *Game of clones: Science is immortalizing Argentina's top polo horses*. Recuperado de <https://www.washingtonpost.com/world/2023/12/29/horse-cloning-polo-argentina/>
10. Clarín. (2013). *Silvina Luna, la primera yegua clonada en el país*. Recuperado de [https://www.clarin.com/sociedad/silvina\\_luna-primera-yegua-clonada\\_0HksEdfZovXI.html](https://www.clarin.com/sociedad/silvina_luna-primera-yegua-clonada_0HksEdfZovXI.html)
11. Infobae. (2023). Clonación, la industria exclusiva en la que Argentina es potencia y el dilema: ¿se podrán duplicar humanos?. Recuperado de <https://www.infobae.com/realidad-aumentada/2023/07/31/clonacion-la-industria-exclusiva-en-la-que-argentina-es-potencia-y-el-dilema-se-podran-duplicar-humanos/>

12. Losinno, C. (2018). Bases para el análisis y mejora de la eficiencia reproductiva en biotecnologías aplicadas a equinos [Documento interno citado en Carrillo Sánchez, 2021].
13. Netessine, S., & Shumsky, R. (2002). Introduction to the theory and practice of yield management. *INFORMS Transactions on Education*, 3(1), 34–44.
14. Aguiar, F. (2004). Teoría de la decisión e incertidumbre: modelos normativos y descriptivos. *Empiria: Revista de Metodología de Ciencias Sociales*, (8), 139–160.
15. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. <https://doi.org/10.1007/BF00116251>
16. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
17. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
18. Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91. <https://doi.org/10.2307/2975974>
19. Adaptado de Regalado, M. (2023). *El proceso de clonación: Los pasos para clonar un caballo* [Infografía]. Infobae. <https://www.infobae.com/realidad-aumentada/2023/07/31/clonacion-la-industria-exclusiva-en-la-que-argentina-es-potencia-y-el-dilema-se-podran-duplicar-humanos/>

## Apéndice A. El proceso de clonación equina: Pasos para clonar un caballo



**Fuente:** Adaptado de Regalado, M. (2023). *El proceso de clonación: Los pasos para clonar un caballo* [Infografía]. Infobae. <https://www.infobae.com/realidad-aumentada/2023/07/31/clonacion-la-industria-exclusiva-en-la-que-argentina-es-potencia-y-el-dilema-se-podran-duplicar-humanos/>