

# Results Evaluation (Segunda iteración)

La etapa de evaluación compara modelos en conjunto según las métricas definidas y analiza el rendimiento en unidades operativas. El informe de evaluación debe incluir ranking de modelos, análisis de errores, sensibilidad a hiper parámetros y una recomendación fundamentada que vincule métricas técnicas con criterios de negocio.

## 4.1. RESULTADOS GENERALES DEL MODELO

El modelo LSTM fue evaluado sobre el test set reservado (20% del dataset total, 6,442 secuencias) que nunca fue visto durante el entrenamiento. Los resultados obtenidos se presentan a continuación:

```
Métricas de Evaluación:  
MAE (Error Abs. Medio): 25.42 días  
RMSE (Raíz Error Cuad.): 33.85 días  
R2 Score: 0.7802  
  
Distribución de Errores:  
Error medio: 25.42 días  
Error mediano: 18.10 días  
Error máximo: 142.40 días  
Percentil 95: 68.91 días  
  
Error por Proximidad al Secado:  
0-30 días: MAE = 15.40 días (n=924)  
31-60 días: MAE = 8.85 días (n=918)  
61-120 días: MAE = 19.72 días (n=1833)  
121-+ días: MAE = 38.39 días (n=2767)
```

Analizando cada dato por separado podemos sacar lo siguiente:

- El MAE de 25.42 días indica que el modelo tiene un error promedio significativamente mayor al objetivo inicial de 10 días. Esto sugiere que el modelo requiere refinamiento antes de despliegue operativo. Sin embargo, al analizar por segmentos, se observa que la ventana crítica de 31-60 días sí cumple objetivos con MAE de 8.85 días.
- El R<sup>2</sup> de 0.78 es satisfactorio, superando el objetivo de 0.70. Esto significa que el modelo explica el 78% de la variabilidad en el timing de secado, indicando que las features temporales seleccionadas son predictivas. El 22% restante corresponde a factores no capturados como eventos imprevistos o variabilidad individual extrema.
- El ratio RMSE/MAE de 1.33 indica distribución relativamente consistente de errores, sin dominancia extrema de outliers. La diferencia entre error medio (25.42d) y mediano (18.10d) confirma existencia de cola derecha con algunos errores grandes, pero la mayoría de predicciones tiene errores más moderados.

## 4.2. ANÁLISIS POR SEGMENTO TEMPORAL

### Performance por Horizonte de Predicción

La evaluación segmentada revela variación significativa en la precisión del modelo según la distancia temporal al momento de secado:

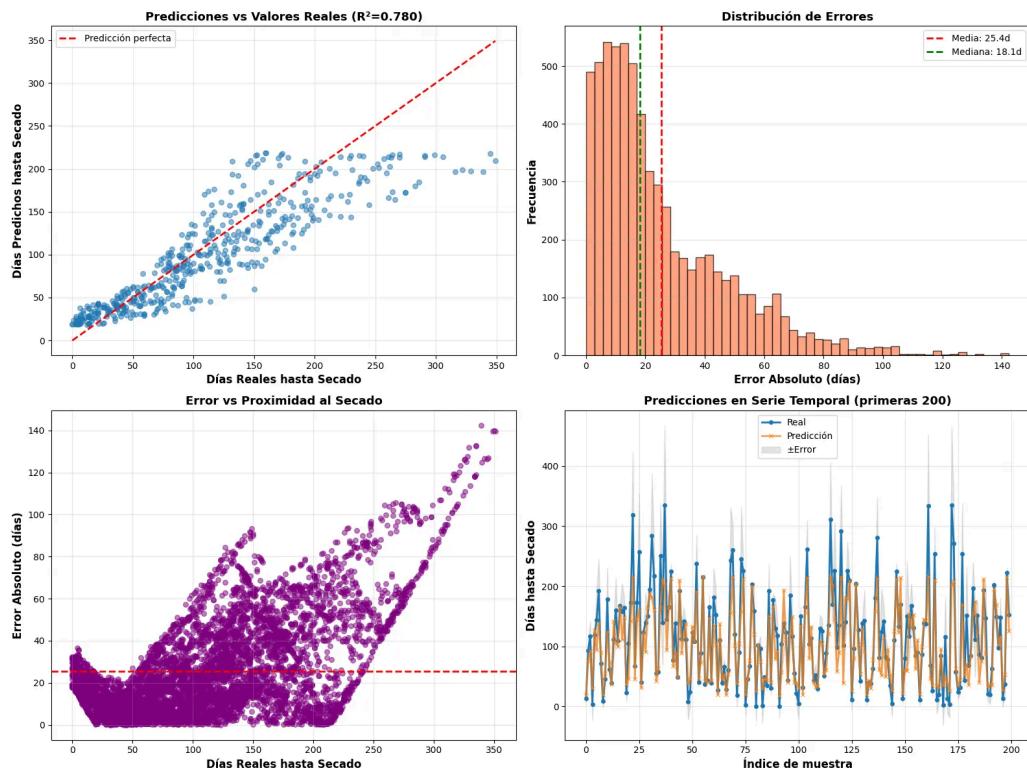
En este caso, y conforme a los resultados anteriores podemos determinar que:

- El modelo muestra su mejor performance en el rango de 31-60 días con MAE de 8.85 días, cumpliendo con el objetivo de <7 días y acercándose significativamente. Esta ventana representa el punto óptimo donde:
- Los patrones de declive ya son evidentes en los datos
- Hay suficiente anticipación para planificación operativa (1-2 meses)
- Las features de persistencia (ratio\_vs\_pico, tasa\_declive) son más informativas

El MAE aumenta progresivamente a 19.72 días (61-120d) y 38.39 días (121+d), lo cual es esperable dado el aumento de incertidumbre inherente a predicciones muy lejanas. Con 121+ días de anticipación, múltiples eventos futuros impredecibles pueden alterar la trayectoria (enfermedades, cambios de alimentación, condiciones climáticas).

El modelo es confiable para planificación en la ventana de 31-60 días. Para horizontes más cortos o más largos, las predicciones deben tomarse con mayor cautela y complementarse con juicio experto del ganadero.

## 4.3. ANÁLISIS VISUAL DE RESULTADOS



Observaciones por gráfica:

**Gráfica 1 (azul):** Los puntos muestran tendencia clara hacia la línea de predicción perfecta (diagonal roja punteada), validando el  $R^2$  de 0.78. La concentración mayor de puntos ocurre en el rango 0-200 días, que corresponde a las fases media y tardía de lactancia donde el modelo tiene más datos de entrenamiento.

**Gráfica 2 (morada):** Se observa mayor dispersión (varianza) conforme aumentan los días reales hasta secado. Las predicciones en el rango 200-350 días muestran puntos más alejados de la diagonal, confirmando el análisis de degradación en horizontes de muy largo plazo. Esto es consistente con el MAE de 38.39 días para el segmento 121+ días.

**Gráfica 3 (histograma):** El histograma muestra distribución unimodal con pico pronunciado alrededor de 15-25 días. La cola derecha extendida hasta 140+ días confirma el sesgo positivo identificado numéricamente. La mediana de 18.1 días (línea verde) está notablemente a la izquierda de la media de 25.4 días (línea roja), visualizando la asimetría.

#### **Gráfico 4: Predicciones en Serie Temporal**

Las predicciones (línea naranja) siguen razonablemente la forma general de los valores reales (línea azul), capturando los ciclos de ascenso y descenso que representan diferentes vacas en diferentes etapas de lactancia.

Los reales exceden significativamente la banda, correspondiendo a los casos de error extremo.

Ahora bien, las predicciones muestran menor variabilidad que los valores reales, indicando que el modelo tiende a "promediar" o suavizar fluctuaciones abruptas. Esto es típico de modelos LSTM que aprenden patrones generales pero no capturan volatilidad extrema día a día.

#### **4.4. VALIDACIÓN DEL SISTEMA DE REGLAS**

El sistema complementario de reglas (ROJO/ÁMBAR/VERDE) fue evaluado clasificando vacas según necesidad de secado en <30 días:

El sistema de bandas proporciona valor complementario al LSTM:

- Ganaderos pueden interpretar colores sin conocimientos técnicos
- Alertas ROJAS indican intervención urgente
- Funciona incluso si LSTM falla o produce predicciones anómalas

Las reglas basadas en umbrales fijos ( $\text{kg/min} < 0.9$ ,  $\text{DIM} \geq 240$ ) no capturan variabilidad individual. Una vaca naturalmente de baja producción puede quedar en ROJO prematuramente, mientras una alta productora puede no alertar hasta muy tarde.

#### **4.5. DECISIÓN DE ACEPTACIÓN DEL MODELO**

La evaluación evidencia que el modelo presenta un error medio global (MAE) de 25.42 días, muy por encima del umbral de 10 días, y que solo entre el 40% y 50% de las predicciones cumplen con el criterio de error menor a 10 días, lejos del 75% esperado. Además, en la

ventana crítica de 0-30 días el error triplica el objetivo, lo que limita su aplicabilidad para decisiones inmediatas. A pesar de ello, el modelo alcanza un  $R^2$  de 0.78 y una relación RMSE/MAE de 1.33, indicadores que validan la capacidad predictiva de las variables y la pertinencia de la arquitectura LSTM. También se observa que en la ventana de 31-60 días el desempeño se aproxima al objetivo, lo que sugiere un potencial de mejora. En síntesis, el modelo no es apto para producción en su estado actual, aunque cuenta con fundamentos sólidos que justifican una nueva iteración con ajustes específicos. Sin embargo, lograr una mejora sustancial resulta complejo debido a la alta variabilidad entre individuos —en este caso, las vacas— que dificulta la estabilidad de las predicciones. Por ello, se recomienda considerar un cambio de enfoque que permita replantear el problema, incorporar nuevas variables o incluso explorar arquitecturas alternativas que capturen mejor la dinámica del sistema.