

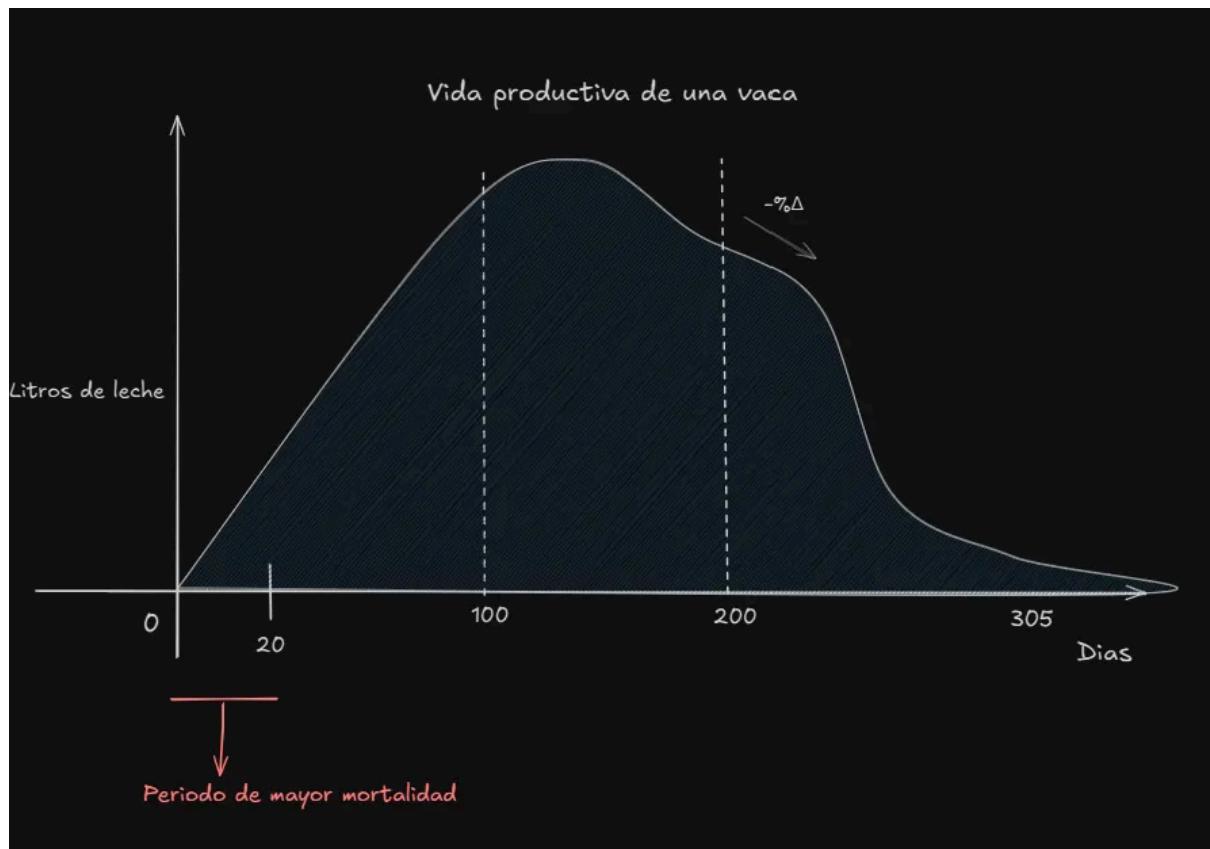
Modeling Technique (Segunda iteración)

Recordando nuestros business objectives queremos detectar incrementos/decrementos en la producción para identificar **vacas secas, no secas o en monitoreo** y así apoyar decisiones veterinarias. En la fase de Modeling Technique se proponen y comparan varios modelos para cumplir nuestros objetivos

Se realizaron **tres iteraciones**, siendo la última la que mejor alineó métricas técnicas y metas de negocio. A continuación, se muestran todas las fases de la parte de modelado dividida por fases

NOTA: En la fase de modelado de la segunda iteración del proyecto, implementamos un enfoque temporal siguiendo las recomendaciones del Prof. Aurelio.

Transformamos nuestro dataset en una serie temporal y concentrarnos específicamente en el último tercio de la lactancia de vacas Holstein (días 200-305), período crítico donde analizamos patrones de declive productivo y tasas de cambio en la producción lechera.



1.1. Propuesta de la primera iteración

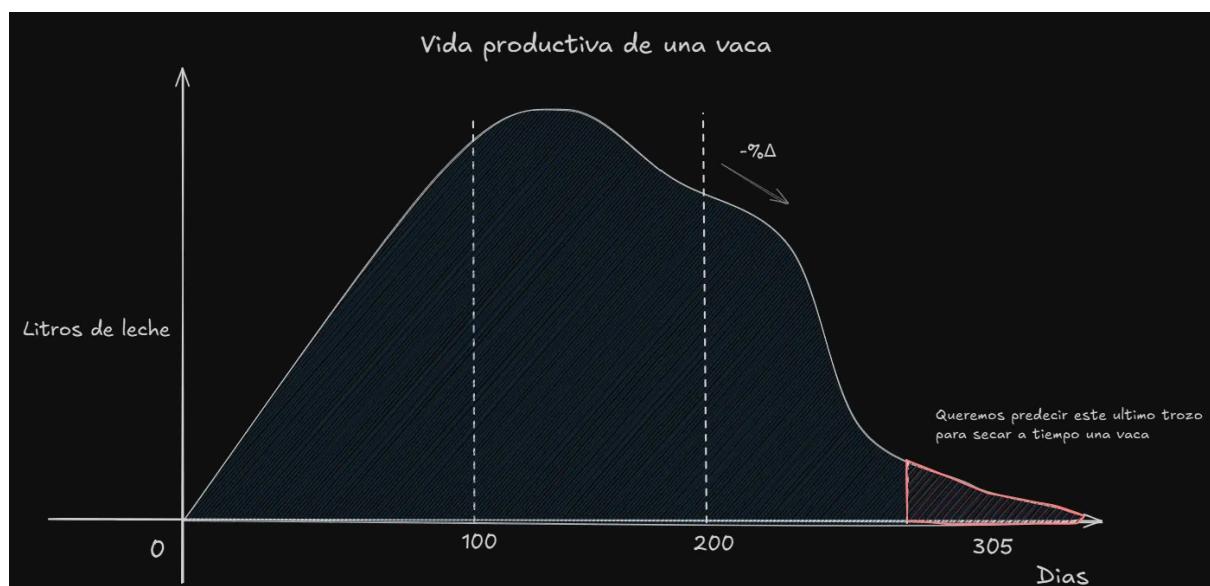
En esta segunda iteración del proyecto implementamos un sistema integral de análisis predictivo para gestión de lactancias en las vacas Holstein. Manejamos nuestro dataset

como una serie temporal, donde la producción del día de hoy, se relaciona a la de ayer. Todo esto siendo procesada mediante una LSTM, permitiendo:

Este enfoque nos permite:

- Identificar **señales tempranas** del fin natural de la lactación
- Modelar la **curva de declive** característica del último tercio y dar una alerta temprana para el periodo de secado de la vaca
- Calcular **métricas de persistencia** y eficiencia residual
- Desarrollar un **sistema predictivo** para el momento óptimo de secado

El sistema permite anticipar de 15-30 días el momento óptimo de secado, optimizando recursos y maximizando la eficiencia del hato



Target: Días de Leche (DEL) en kg por ordeño

Tipo de problema: Regresión logística (variable continua)

1.2. Proceso de la primera iteración

Basándonos en las recomendaciones del Prof. Aurelio y en el análisis de las limitaciones de la primera iteración, rediseñamos completamente nuestro enfoque. Reconocimos que el problema de predecir el momento óptimo de secado es fundamentalmente temporal. La producción de leche no es un valor aislado, sino parte de una secuencia donde cada medición se relaciona con las anteriores y proporciona contexto para interpretar las futuras.

Esta comprensión nos llevó a replantear el problema como una tarea de predicción de series temporales. En lugar de preguntar "¿esta vaca está produciendo poco hoy?", ahora preguntamos "¿cuál es la trayectoria de producción de esta vaca y cuándo alcanzará el punto donde el secado es óptimo?". Este cambio de paradigma requirió transformar completamente nuestra estructura de datos y seleccionar técnicas de modelado apropiadas para capturar dependencias temporales.

1.3. Evolución desde la primera iteración

Limitaciones de la Iteración 1:

- Cada ordeño tratado como evento independiente, sin contexto temporal
- No capturaba patrones de declive ni tendencias
- El ajuste de los datos era muy pequeño
- Sin capacidad predictiva anticipatoria

1.4. Criterios de Selección:

Modelo Seleccionado: LSTM

Criterio 1 - Naturaleza temporal del problema:

- LSTM fue diseñado específicamente para secuencias con dependencias de largo plazo.
- Incorpora memoria explícita que retiene información durante muchos timesteps (30+ días en nuestro caso).

Criterio 2 - Mecanismo de puertas:

- Descartar información irrelevante (eventos anómalos pasados que ya no importan).
- Incorpora nueva información significativa (inicio de tendencia de declive sostenido).
- Decide qué información del estado de memoria usar para la predicción actual.
- **Ventaja para nuestro problema:** Filtra variación diaria irrelevante mientras retiene señales de tendencia de 10-15 días..

Criterio 3- Captura de no linealidad:

- Maneja la curva no lineal de lactancia (ascenso rápido → pico día 60-100 → declive gradual acelerado).

Criterio 4 - Adecuación al volumen de datos:

- A diferencia de Transformers, LSTM funciona efectivamente con datasets moderados.
- A diferencia también de una MLP, considera temporalidad
- Balance óptimo entre capacidad de aprendizaje y requerimientos de datos.

1.5. Supuestos Generales

- **Dependencia de observaciones:** Cada registro de ordeño es dependiente. La producción de ayer se relaciona con la actual.
- **Completitud de datos:** Los registros faltantes fueron eliminados y no introducen sesgo sistemático.
- **Representatividad:** Los datos históricos son representativos del comportamiento futuro de producción.

1.6. Supuestos por Modelo

Supuesto	Descripción	Implicación/Mitigación
Continuidad temporal	Secuencias son continuas con ordeño diario regular	Gaps <2 días: forward-fill; Gaps >2 días: exclusión del ciclo
Normalización [0,1]	Todas las features normalizadas con MinMaxScaler	Valores fuera de rango se saturan; requiere monitoreo continuo
Capacidad de memoria	128 unidades LSTM suficientes para 30 días × 50+ features	Validado empíricamente; aumentar unidades no mejora significativamente
Dropout como regularizador	Dropout 30% (LSTM) y 20% (Dense) previene overfitting	Trade-off: capacidad vs. generalización dado dataset pequeño
Predicción puntual	Modelo genera estimación única, no intervalos de confianza	Simplifica arquitectura; considerar Monte Carlo Dropout futuro