

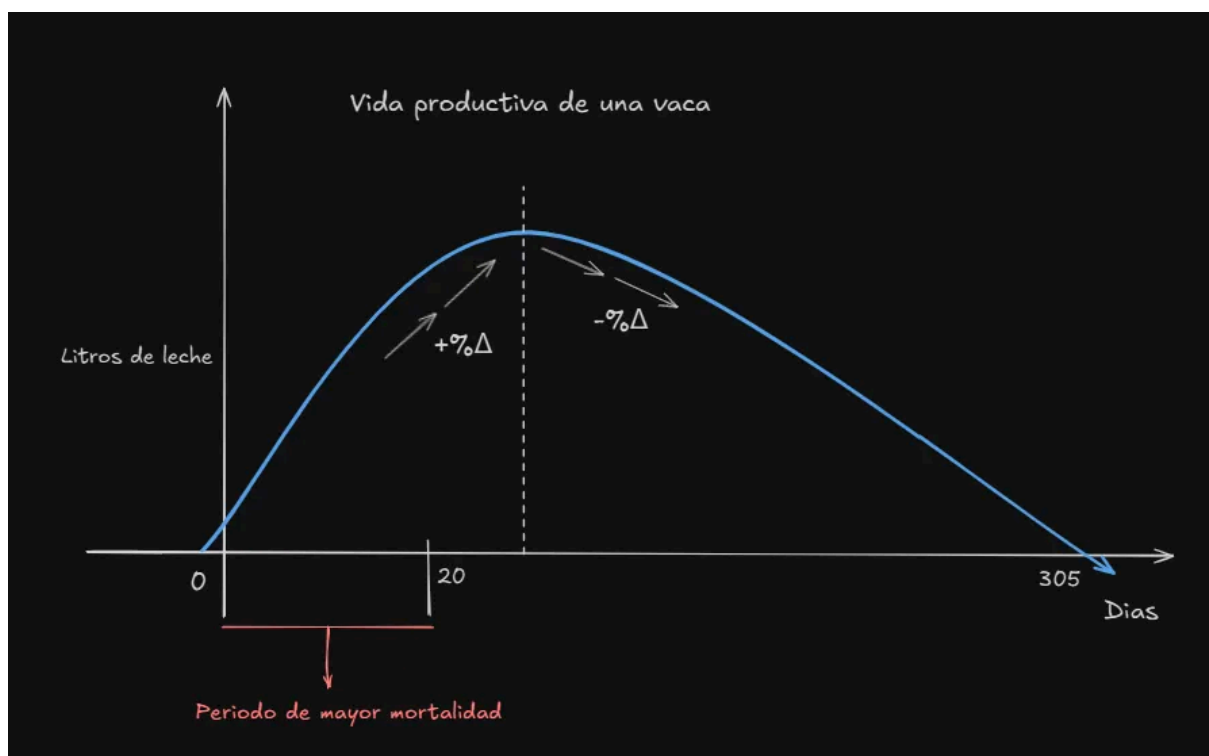
Modeling Technique (Primera iteración)

Recordando nuestros business objective 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.

Colab con la primera iteración: [RF_CV_Datos de ordeña - Todas las vacas](#)

*NOTA: Se desarrolló un modelo predictivo para estimar **Días de Leche (DEL, kg)** por ordeño por vaca; el mejor modelo fue **Random Forest** explicando **71.06%** de la varianza. Sin embargo, el poder predictivo y el objetivo de minería de datos no se cumplió.*



1.1. Propuesta de la primera iteración

La idea en la que se basó esta primera iteración del modelo es en identificar decaimientos en la curva de lactancia para, cuando se viera un cambio negativo ($-\Delta$ en producción) podamos clasificar una VACA SECA y NO SECA.

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

Tipo de problema: Regresión logística (variable continua)

1.2. Proceso de la primera iteración

Técnicas implementadas

Algoritmo	Categoría	Justificación
Regresión Lineal	Modelo lineal baseline	Modelo simple e interpretable para establecer línea base
Ridge Regression	Modelo lineal regularizado	Manejo de multicolinealidad entre variables
Lasso Regression	Modelo lineal regularizado	Selección automática de features relevantes
Random Forest	Ensemble basado en árboles	Captura relaciones no lineales e interacciones complejas

El abanico de técnicas implementadas incluyó **algoritmos lineales** como Regresión Lineal, Ridge y Lasso, junto con **algoritmos basados en árboles** como Random Forest. Esta selección diversa permitió evaluar diferentes enfoques para capturar las relaciones subyacentes en los datos, desde modelos simples e interpretables hasta métodos más complejos capaces de capturar interacciones no lineales.

La elección de estas técnicas se basó en criterios técnicos bien definidos. Dada la **naturaleza del problema** como una tarea de regresión con una variable continua (Producción en kg), era esencial seleccionar algoritmos apropiados para este tipo de predicción. El **tamaño del dataset**, aproximadamente 4,000 registros después del proceso de limpieza, no permite el uso de algoritmos más complejos o de técnicas de Deep Learning.

La naturaleza mixta de las **características**, que incluían tanto variables numéricas como categóricas, requería algoritmos capaces de manejar esta diversidad de tipos de datos.

1.3. Criterios de Selección:

- **Naturaleza del problema:** Regresión con variable continua (Producción en kg)
- **Tamaño del dataset:** ~4,000 registros después de limpieza (tamaño moderado)
- **Tipos de características:** Mezcla de variables numéricas (DEL, días de lactancia) y categóricas (Como ID vaca)

- **Complejidad progresiva:** Comenzar con modelos simples (baseline) y progresar hacia modelos más complejos
- **Interpretabilidad:** Equilibrio entre poder predictivo y explicabilidad para stakeholders veterinarios.

Con este primer enfoque pretendemos progresar a modelos más complejos para capturar relaciones no lineales y ver si es que este primer approach es lo suficientemente fuerte para solucionar el problema

1.4. Supuestos Generales

- **Independencia de observaciones:** Cada registro de ordeño es independiente (aunque en realidad existen dependencias temporales dentro de cada vaca)
- **Complejidad de datos:** Los registros faltantes fueron eliminados y no introducen sesgo sistemático
- **Representatividad:** Los datos históricos son representativos del comportamiento futuro de producción

1.5. Supuestos por Modelo

- **Regresión Lineal/Ridge/Lasso:**
 - Relación lineal entre predictores y target
 - Homocedasticidad de residuos
 - Distribución normal de errores
- **Random Forest:**
 - No requiere supuestos de linealidad o distribución
 - Suficientes observaciones por nodo para splits significativos
 - Complejidad o Sencillez para ser un algoritmo de ML