

Features engineering (Tercera iteración)

Dataset de Fichas de Vaca (HATO)

Durante la fase de feature engineering se genera información reproductiva esencial para el análisis del comportamiento productivo de las vacas. Para ello, se implementa una función que asigna dos nuevas variables:

1. Estado Reproductivo
2. Número acumulado de inseminaciones

Esta transformación se realiza porque el dataset original solo registra eventos individuales (como partos, inseminaciones o secados), pero no proporciona un estado reproductivo actualizado para cada día ni un historial acumulado de inseminaciones. Para que los modelos puedan interpretar la evolución reproductiva de cada vaca, es necesario reconstruir esta información de forma secuencial.

Dataset de Registros de ordeño (HATO)

La agregación diaria y la creación de características derivadas son pasos esenciales para transformar los registros individuales de ordeño en información útil y estable para el análisis. Los datos originales presentan variabilidad natural debido a diferencias entre ordeños, intervalos irregulares, efectos circadianos y ruido de medición, lo que dificulta la detección de tendencias y patrones reales. Al consolidar la información por día y por vaca, se obtiene un conjunto de métricas más consistente que refleja con mayor precisión el rendimiento productivo diario.

Esta transformación también es coherente con el manejo ganadero, ya que la producción láctea y los fenómenos fisiológicos relevantes operan en escalas diarias. Las decisiones de nutrición, salud y gestión se basan en valores acumulados por día y no en registros individuales, por lo que el análisis debe alinearse con esta lógica.

Posteriormente, se generan características derivadas como la producción por minuto y el intervalo promedio entre ordeños, las cuales permiten evaluar eficiencia, detectar irregularidades y comparar comportamientos entre animales. También se calculan las lactaciones y los días en leche (DEL), lo que contextualiza cada registro dentro del ciclo productivo de la vaca y permite interpretar su posición en la curva de lactancia. Finalmente, el uso de ventanas móviles suaviza las fluctuaciones cotidianas y facilita la identificación de tendencias recientes.

Estas transformaciones permiten construir un dataset robusto, interpretable y adecuado para análisis descriptivos, predictivos y de caracterización del hato.