

Descripción del dataset final (Segunda iteración)

El dataset final de esta iteración contiene registros a nivel de evento de ordeño para las producciones de las vacas, donde cada fila representa una sesión de ordeño en un momento específico. Aunque la estructura exacta puede variar según las columnas presentes tras la carga, de forma general el dataset incluye:

- Columnas temporales y de identificación
 - datetime: marca de tiempo del inicio del ordeño (en formato datetime).
 - (En caso de estar presente) Número de ordeño: identificador de la sesión dentro del día.
- Variables fisiológicas y productivas por cuarto
 - DI - media de flujos (kg/min)
 - DD - media de flujos (kg/min)
 - TI - media de flujos (kg/min)
 - TD - media de flujos (kg/min)
 - DI - sangre (ppm), DD - sangre (ppm), etc.
 - DI - conductividad (ms/cm), DD - conductividad (ms/cm), etc.
 - DI - producciones (kg), DD - producciones (kg), etc.
- Variables agregadas de producción total
 - Producción (kg): producción total de leche en la sesión de ordeño.
- Variables derivadas logarítmicas
 - log_produccion (o nombres equivalentes que derivan de Producción (kg)).
 - log_DI_sangre, log_DD_sangre, etc.
 - Variables de la forma log_DI_flows, log_DD_flows, o equivalentes según la lógica definida en el pipeline.

Finalmente entonces podremos utilizar modelos temporales para esta problemática como LSTM, ARIMA/SARIMA.