

Transformaciones (Segunda iteración)

Dataset de Registros de Ordeño

La principal transformación se enfoca en la gestión de la fecha y hora del ordeño. En el dataset original, la información temporal se encuentra en una única columna de texto (Hora de inicio).

Este formato no es directamente utilizable para análisis de series de tiempo, por lo que se realiza el siguiente proceso:

1. Normalización del formato de hora
 - Se reemplazan las cadenas "a. m." y "p. m." por los formatos estándar "AM" y "PM".
2. Conversión a tipo datetime

Además de la gestión temporal, se aplica un renombramiento de columnas para estandarizar la nomenclatura de las mediciones por cuarto de la ubre. De:

- DI, DD, TI, TD
- DI.1, DD.1, TI.1, TD.1
- DI.2, DD.2, TI.2, TD.2
- DI.3, DD.3, TI.3, TD.3
- DI.4, DD.4, TI.4, TD.4

se convierten a nombres descriptivos del tipo:

- DI - media de flujos (kg/min)
- DD - media de flujos (kg/min)
- DI - sangre (ppm)
- TI - conductividad (ms/cm)
- TD - producciones (kg)

Esta transformación estandariza el significado de cada columna, mejora la legibilidad del dataset y reduce la ambigüedad al momento de interpretar correlaciones y resultados del modelo.