

Resumen de la Preparación de Datos

Los datasets utilizados para esta fase son las fichas de vacas y los registros de ordeña por hato; ambos requieren transformaciones esenciales para cumplir con los objetivos de minería de datos establecidos en el [Documento de Alineación del Proyecto \(DAP\)](#).

Los datos cargados cuentan con un análisis y comprensión previa, los cuales pueden consultarse en el documento [Reporte de Comprensión de Datos \(RCD\)](#) donde se describen a profundidad las limitaciones y características de la información utilizada.

La principal limitación de estos datos es que cada vaca cuenta con su propio archivo CSV. Debido a que el nombre del archivo contiene el identificador de la vaca, es necesario aplicar una estrategia que permita extraer dicho identificador e incorporarlo como una nueva columna dentro del dataset. Esto facilita la construcción de un dataset unificado y completo que integre la información de todas las vacas registradas en el CAETEC.

Considerando los objetivos de minería de datos (Data Mining Goals) y tras la validación realizada por expertos en el área, se concluye que los datos generados y transformados cumplen adecuadamente con los requerimientos establecidos. Por lo tanto, se puede proceder con la siguiente fase del proceso: el modelado.

- **Data Mining Goals**
 - **Objetivo de Negocio 1:** Asegurar que los resultados del proyecto puedan replicarse en otros hatos y/o ranchos.
 - **Objetivo de Minería de Datos**
 - Evaluar consistencia de patrones estables de comportamiento reproductivo en diferentes subconjuntos del hato.
 - Determinar replicación de los hallazgos de las variables reproductivas entre períodos o grupos de vacas.
 - **Objetivo de negocio 2:** Aportar conocimiento útil sobre las vacas Holstein.
 - **Objetivo de Minería de Datos**
 - Determinar variables que expliquen el estado reproductivo de las vacas Holstein.
 - Generar conocimiento de indicadores del estado reproductivo de las vacas Holstein.
 - **Objetivo de negocio 3:** Respaldar decisiones de manejo y cuidados de la vaca para su secado.
 - **Objetivos de Minería de Datos**
 - Generar información preventiva sobre patrones o condiciones recurrentes que anteceden mejoras o disminuciones en el estado reproductivo.