

Operación	Tipo de dato afectado	Objetivo (técnico)	Objetivo (forense / auditabilidad)	Efecto práctico
Normalización de nombres de columna	Todas	Unificar esquema entre datasets	Repetibilidad: mismo script, mismo resultado	Evita fallos por cambios de nombres
Eliminación/gestión de NaN e infinitos	Numéricas	Estabilidad del modelo y métricas	Evitar resultados no reproducibles por errores de lectura	Sustitución por 0/mediana según regla
Conversión de tipos (casting)	Mixto	Consistencia del entrenamiento	Evitar ambigüedad en interpretación de variables	label int; categóricas string
Regla attack_cat/label	Etiquetas	Integridad del ground truth	Evitar contradicciones en evidencia derivada	Normal→0; ataque→1
Selección de columnas comunes (real vs syn)	Variables	Comparabilidad en validación	Defensa: comparar 'mismas entradas'	Evita que el modelo vea features distintas
One-Hot Encoding (OHE)	Categóricas	Transformar para modelos tabulares	Trazabilidad: mapping estable de categorías	Permite reproducir matriz de features
Lectura por chunks	Todas	Evitar colapsos por memoria	Repetibilidad operativa en entorno estándar	Posibilita procesar datasets muy grandes