

| Dataset | Origen / contexto | Filas reales aprox. (full_clean) | Variables (tras limpieza) | attack_cat (nº clases) | Uso en este TFM | Observaciones |
|---------------------------------|---|-------------------------------------|------------------------------|------------------------------------|---|---|
| CICIDS2017 (MachineLearningCVE) | Tráfico realista (varios días) con múltiples perfiles de ataque | ~2.830.743 | ~80 | 4 (Normal + 5 ataques principales) | Dataset principal (detección binaria + XAI) | Muy adecuado para evaluación binaria; buen equilibrio y riqueza de variables |
| UNSW-NB15 | Tráfico corporativo con taxonomía amplia de ataques | ~2.540.047 | ~45 | 9 (Normal + 8 ataques) | Comparación breve | Multiclas exigente por minorías; útil para 'stress test' del enfoque |
| Kitsune (IoT) | Tráfico IoT con variables agregadas (escenario de baja dimensionalidad) | ~20.438.941 | ~7 | 7 (Normal + 6 ataques) | Comparación breve | Muy grande en filas pero pocas variables; limita explicabilidad 'rica' y generalización |