|  |
| --- |
| **Dataset WDBC (densidad=25, riesgo=50, pureza=proporcion)** |

**1. Identificación general**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **dataset** | **clase** | **train\_original** | **objetivo\_balance** | **estado** | **motivo\_sin\_sinteticas** |
| wdbc | 0 | 285 | 228 | no se sobremuestrea | sin\_faltante(actual>=objetivo) |
| wdbc | 1 | 170 | 228 | sobremuestreada | ok |

**2. Resultados de tamaño y ratios**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **clase** | **muestras\_sinteticas\_generadas** | **total\_original** | **total\_resampled** | **ratio\_original** | **ratio\_resampled** |
| 0 | 0 | 455 | 455 | 0.626374 | 0.626374 |
| 1 | 58 | 455 | 513 | 0.373626 | 0.344595 |

**3. Métricas internas de riesgo/densidad**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **clase** | **n\_candidatas** | **n\_filtradas** | **riesgo\_medio** | **riesgo\_std** | **densidad\_media** | **vecinos\_validos\_promedio** |
| 0 | — | — | — | — | — | — |
| 1 | 43 | 43 | 0.7628 | 0.2124 | 0.0651 | 3.0 |

**4. Parámetros de configuración**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **clase** | **umbral\_densidad** | **umbral\_entropia** | **percentil\_densidad** | **percentil\_riesgo** | **criterio\_pureza** | **tecnica\_sobremuestreo** | **factor\_equilibrio** | **random\_state** | **modo\_espacial** | **timestamp** |
| 0 | — | — | 25.0 | 50.0 | **proporcion** | PCSMOTE | 0.8 | 42 | 2d | 2025-08-20T00:20:16.037738 |
| 1 | 0.0 | — | 25.0 | 50.0 | **proporcion** |  |  |  |  |  |

**Tabla 1 – Datos generales de configuración**

Contiene la información básica de cada dataset, el tipo de muestreo, la técnica aplicada y los parámetros principales de PCSMOTE.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **dataset** | **tipo** | **tecnica** | **densidad** | **riesgo** | **pureza** | **n\_train** |
| wdbc | base | base | – | – | – | 398 |

**Tabla 2 – Detalles del modelo y configuración óptima**

Muestra la cantidad de test, número de features, folds de CV, iteraciones, modelo usado y la mejor configuración de hiperparámetros encontrada.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **n\_test** | **n\_features** | **es\_grande** | **cv\_splits** | **n\_iter** | **modelo** | **mejor\_configuracion** | **cv\_f1\_macro** |
| 171 | 30 | False | 5 | 10 | LogisticRegression | {'classifier\_\_C': np.float64(0.4042872735027331), 'classifier\_\_penalty': 'l2', 'classifier\_\_solver': 'lbfgs'} | 0.967913 |

**Tabla 3 – Resultados de métricas de validación y test**

Presenta los resultados de validación cruzada y test en todas las métricas calculadas.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **cv\_balanced\_accuracy** | **cv\_mcc** | **cv\_cohen\_kappa** | **test\_f1\_macro** | **test\_balanced\_accuracy** | **test\_mcc** | **test\_cohen\_kappa** | **search\_time\_sec** | **n\_jobs\_search** |
| 0.967655 | 0.936618 | 0.935868 | 0.987515 | 0.987515 | 0.975029 | 0.975029 | 5.386 | 4 |