

Concepto	Qué hace	Ventajas (por qué)	Desventajas (por qué)	Cuándo usarlo
<code>shuffle</code>	Mezcla aleatoriamente las filas del dataset.	<ul style="list-style-type: none"> <li>- Evita patrones o sesgos si los datos vienen ordenados.</li> <li>- Mejora la generalización.</li> </ul>	<ul style="list-style-type: none"> <li>- Puede alterar el equilibrio de clases si se usa solo.</li> <li>- Si no se controla, puede dar conjuntos desbalanceados.</li> </ul>	<ul style="list-style-type: none"> <li>- Siempre que prepares datos antes de dividir.</li> <li>- Especialmente importante si los datos vienen ordenados por fecha o clase.</li> </ul>
<code>stratify</code>	Mantiene la proporción de clases al dividir el dataset.	<ul style="list-style-type: none"> <li>- Evita desbalanceo entre entrenamiento y test.</li> <li>- Muy útil en clasificación con clases desiguales.</li> </ul>	<ul style="list-style-type: none"> <li>- No funciona con regresión directa (porque no hay clases).</li> <li>- Puede fallar si alguna clase tiene muy pocos ejemplos.</li> </ul>	<ul style="list-style-type: none"> <li>- En problemas de clasificación.</li> <li>- Cuando hay clases desbalanceadas (unas con muchos datos y otras con pocos).</li> </ul>
<code>shuffle</code> + <code>stratify</code>	Mezcla los datos pero respetando las proporciones de clase.	<ul style="list-style-type: none"> <li>- Combina lo mejor de ambos: aleatoriedad y equilibrio.</li> <li>- Aumenta la robustez del modelo y representatividad del test.</li> </ul>	<ul style="list-style-type: none"> <li>- Más lento computacionalmente si el dataset es muy grande.</li> <li>- Puede ser innecesario en conjuntos perfectamente balanceados.</li> </ul>	<ul style="list-style-type: none"> <li>- Recomendado por defecto en clasificación.</li> <li>- Úsalo en <code>train_test_split</code> o validación cruzada (<code>cross_val_score</code>).</li> </ul>