1. Random Splitting Algorithm

The Random Splitting algorithm shuffles the dataset before dividing it, ensuring that the model is trained on a variety of samples regardless of their original position.

Algorithm: Random Train-Test Split

- **Input:**
  - Dataset D
  - Test size ratio r (e.g., 0.2 for 20% test set)
- **Output:**
  - Training set T_train
  - Testing set T_test

1. **Shuffle** dataset D randomly.
2. **Calculate** split_index = r * size(D).
3. Set T_test = first split_index samples from D.
4. Set T_train = remaining samples from D.
5. **Return** T_train and T_test.

---

2. Stratified Splitting Algorithm

The Stratified Splitting algorithm ensures that the class distribution (proportion of labels) in the original dataset is preserved in both subsets. This is essential for handling imbalanced datasets.

Algorithm: Stratified Train-Test Split

- **Input:**
  - Dataset D with class labels Y
  - Test size ratio r
- **Output:**
  - T_train
  - T_test

1. **For each class** c in Y:
   - a. **Extract** all samples belonging to class c.
   - b. **Shuffle** samples of class c.
   - c. **Split** class c samples into train and test using ratio r.
2. **Combine** all class-wise train samples into T_train.
3. **Combine** all class-wise test samples into T_test.
4. **Return** T_train and T_test.

---

3. Sequential Splitting Algorithm

The Sequential Splitting algorithm divides the data without shuffling. This maintains the chronological or logical order of the data, which is critical for time-series analysis.

Algorithm: Sequential Train-Test Split

- **Input:**
  - Dataset D
  - Test size ratio r
- **Output:**
  - T_train
  - T_test

1. **Calculate** split_index = (1 - r) * size(D).
2. Set T_train = first portion of D up to split_index.
3. Set T_test = remaining portion of D starting from split_index.
4. **Return** T_train and T_test.

Summary of Splitting Methods

Algorithm	Shuffling	Order Preservation	Primary Use Case
Random Splitting	Yes	No	General purpose, IID data
Stratified Splitting	Yes (within class)	No	Imbalanced datasets
Sequential Splitting	No	Yes	Time-series, sequence data