**Imbalance Data Technique**: A dataset is imbalanced when the number of observations in one class is significantly higher than in other classes.

Imbalanced datasets are **very common in classification problems**, especially in areas like:

* Fraud detection
* Disease diagnosis
* Spam detection

1. Oversampling (Add more minority sample):

#### **Techniques:**

**Random Undersampling** Randomly drops majority class samples.  
**Tomek Links** Removes ambiguous points near the decision boundary.  
 **Cluster Centroids** Use clustering (e.g., KMeans) to reduce the majority class to its centroid representation.

1. Undersampling (Remove majority sample):