

Stratified Sampling



Random Sampling

Original Dataset (70% A, 30% B)



Train



83% A 😞

Val



50% A 😞

Test



0% A 🤯



Stratified Sampling

Original Dataset (70% A, 30% B)



Train



70% A ✓

Val



70% A ✓

Test



70% A ✓

Why Stratified Sampling?



Critical for imbalanced datasets



Prevents biased evaluation



Reduces variance in estimates