



# Train-Test Ratio

# Train-Test Ratio

Original Dataset

Train Data - 70%

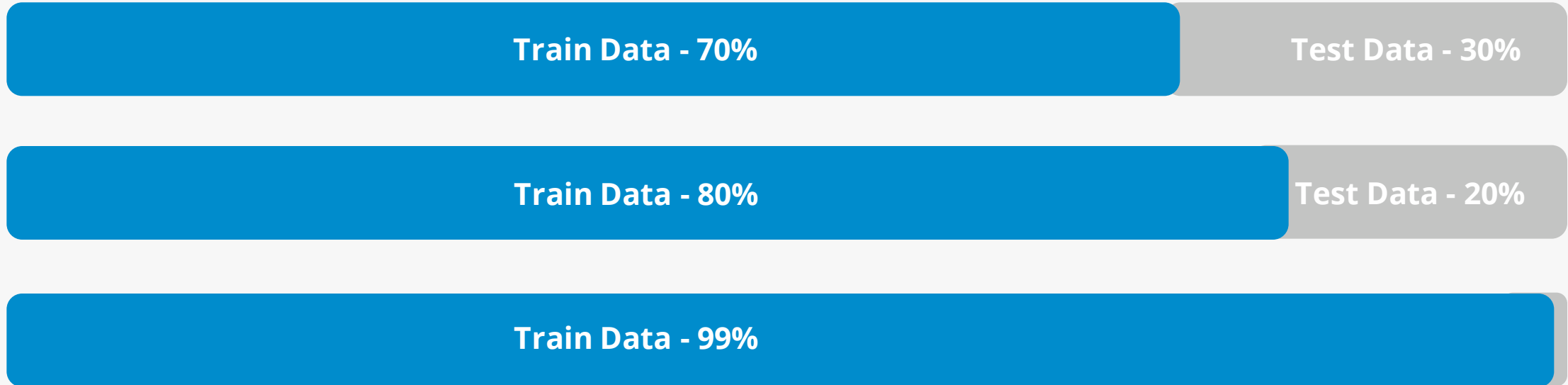
Test Data - 30%

Train Data - 80%

Test Data - 20%

Train Data - 99%

# Train-Test Ratio



Train data is kept higher to ensure sufficient data is available to make insights and build the prediction system.

# Train-Test Ratio

Train Data - 70%

Test Data - 30%

Train Data - 80%

Test Data - 20%

Train Data - 99%

Test data should have sufficient samples to reduce the impact of sample bias.

# Train-Test Ratio: 500 rows

**Original Dataset**

**Train Data - 70%**

**Test Data - 30%**

# Train-Test Ratio: 10,000 rows

**Original Dataset**

**Train Data - 80%**

8000 rows

**Test Data - 20%**

2000 rows

# Train-Test Ratio: 100,000 rows

Original Dataset

Train Data - 90%

10%

90000 rows

10000 rows

# Train-Test Ratio: 100 million rows

Original Dataset

Train Data - 99%

99000000 rows

10000000 rows





# Limitations of Train-Test Split

# Limitations of Train-Test Split

- **Difference in Evaluation**

1. Highly dependent on specific random split of data
2. Different splits lead to different evaluation performance



# Limitations of Train-Test Split

- **Less Efficient Data Utilization**

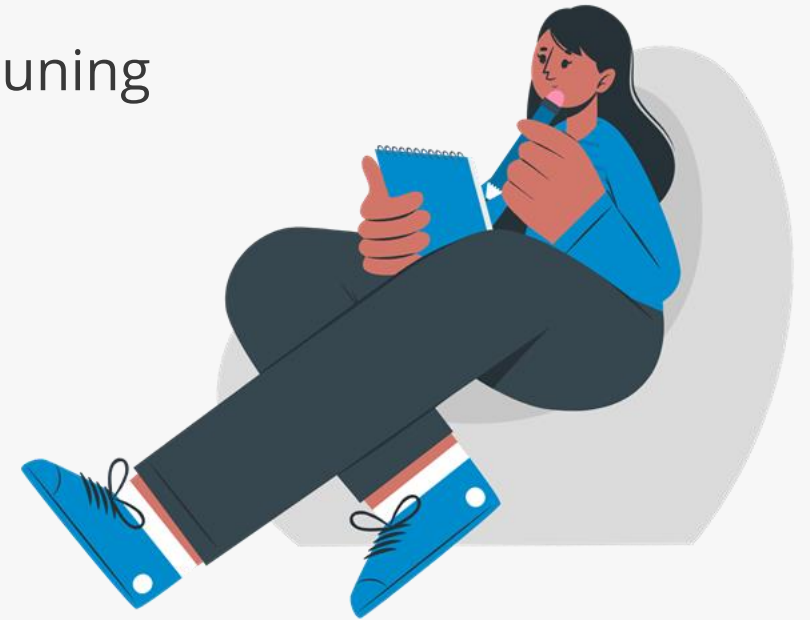
1. A portion of data is reserved for testing
2. Amount of data available for training may not be sufficient



# Limitations of Train-Test Split

- **Limited Information for Model Selection**

1. Train-Test Split provides only one evaluation of mode's performance
2. Might not be reliable especially when it comes to tuning hyperparameters





# Cross Validation