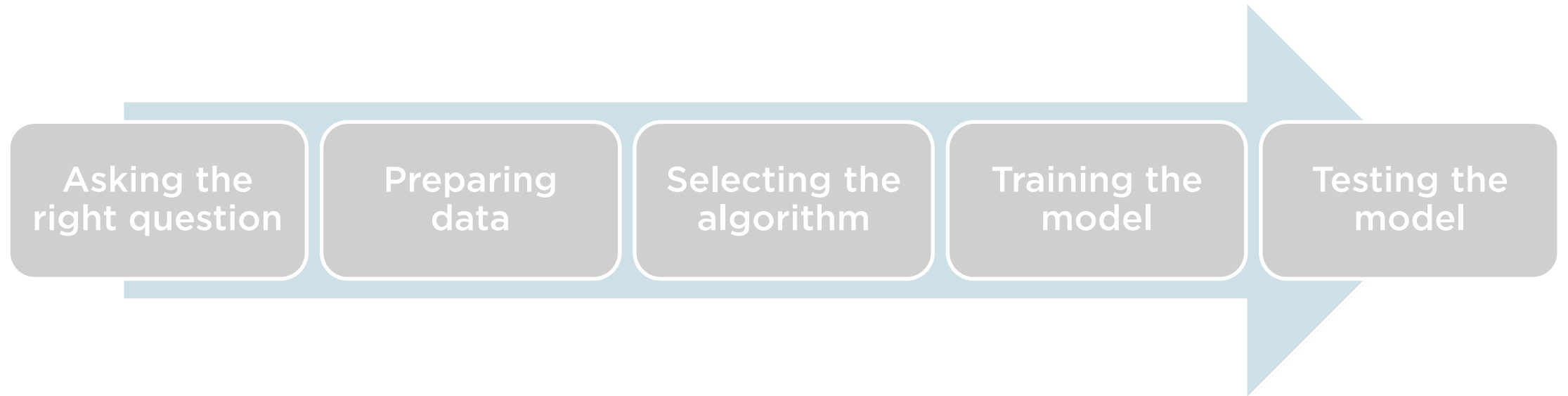# Training the Model

Jerry Kurata
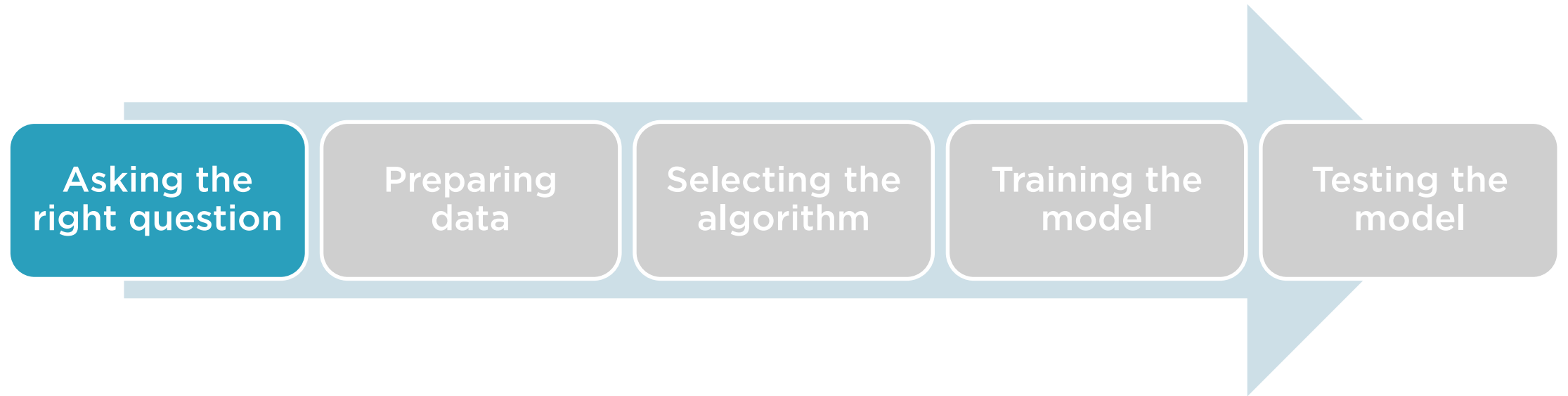CONSULTANT

@jerrykur   www.insteptech.com
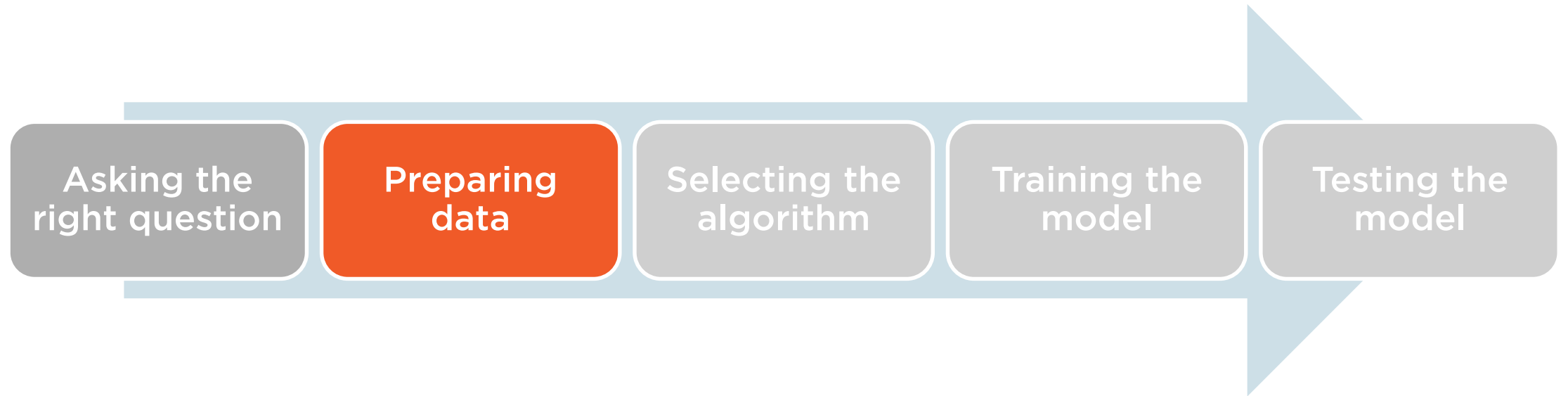
Machine Learning Workflow
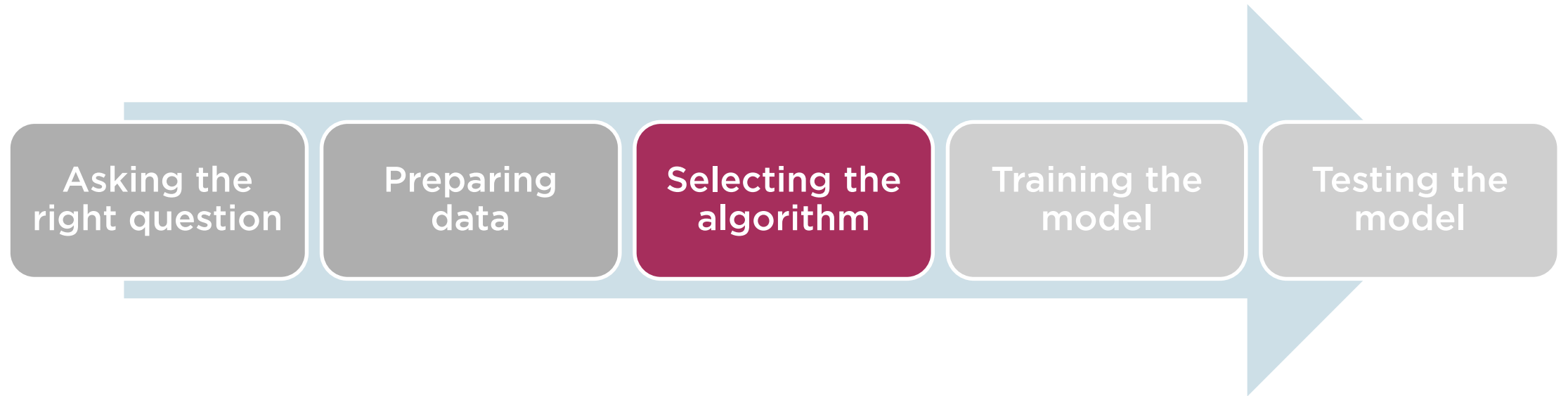
Machine Learning Workflow

Machine Learning Workflow

Asking the right question → **Preparing data** → Selecting the algorithm → Training the model → Testing the model

Machine Learning Workflow

Asking the right question → Preparing data → **Selecting the algorithm** → Training the model → Testing the model

Machine Learning Workflow

Asking the right question → Preparing data → Selecting the algorithm → **Training the model** → Testing the model
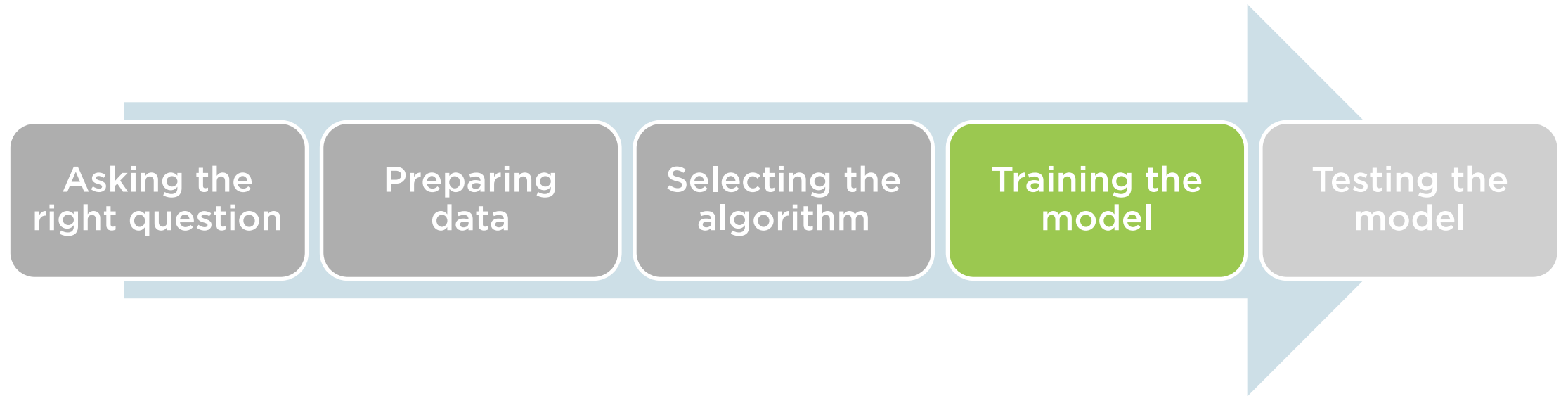
Overview

Understand the training process

Caret package

Train algorithm with DOT data

Machine Learning Training

Letting specific data teach a Machine Learning algorithm to create a specific forecast model.

Machine Learning Training

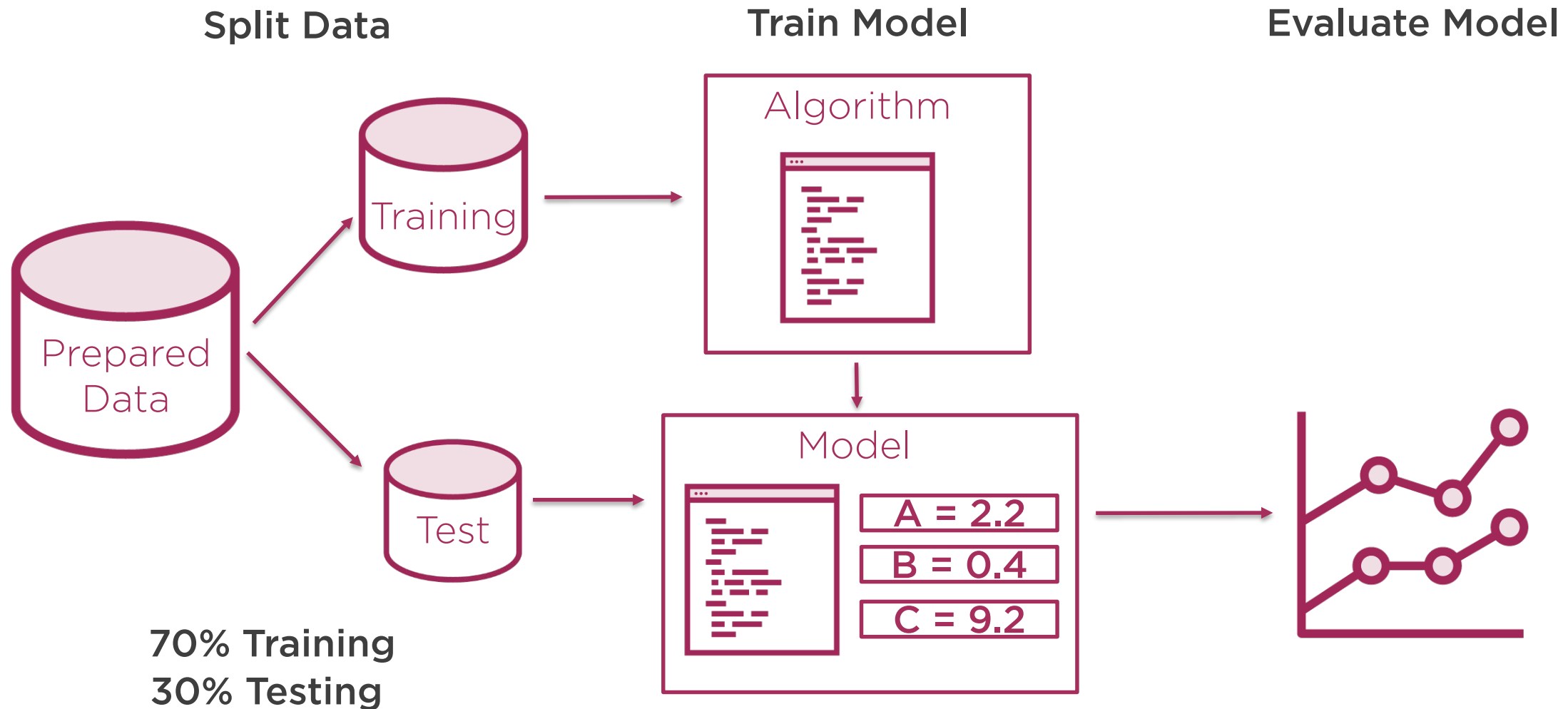Letting specific data teach a Machine Learning algorithm to create a specific forecast model.

Why retrain?

New data => better predictions

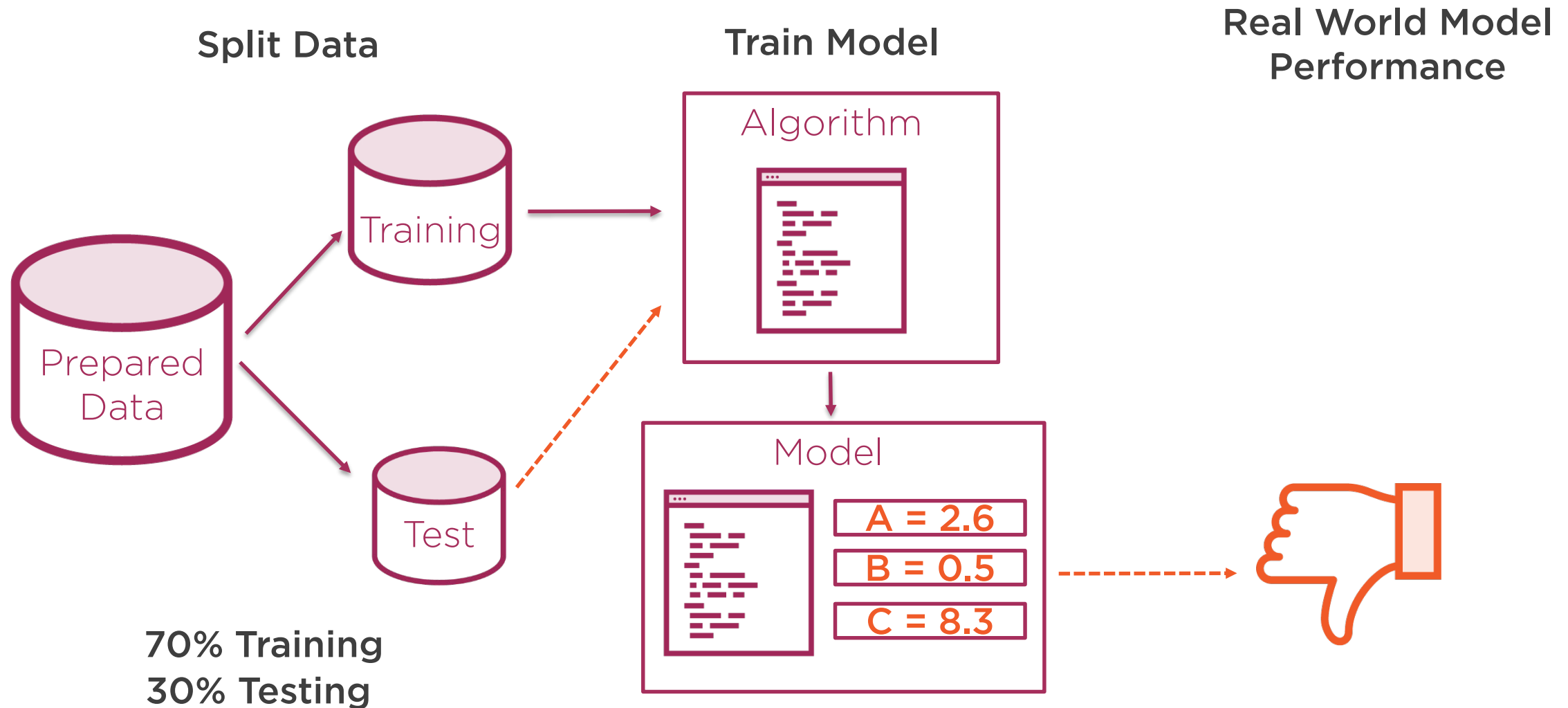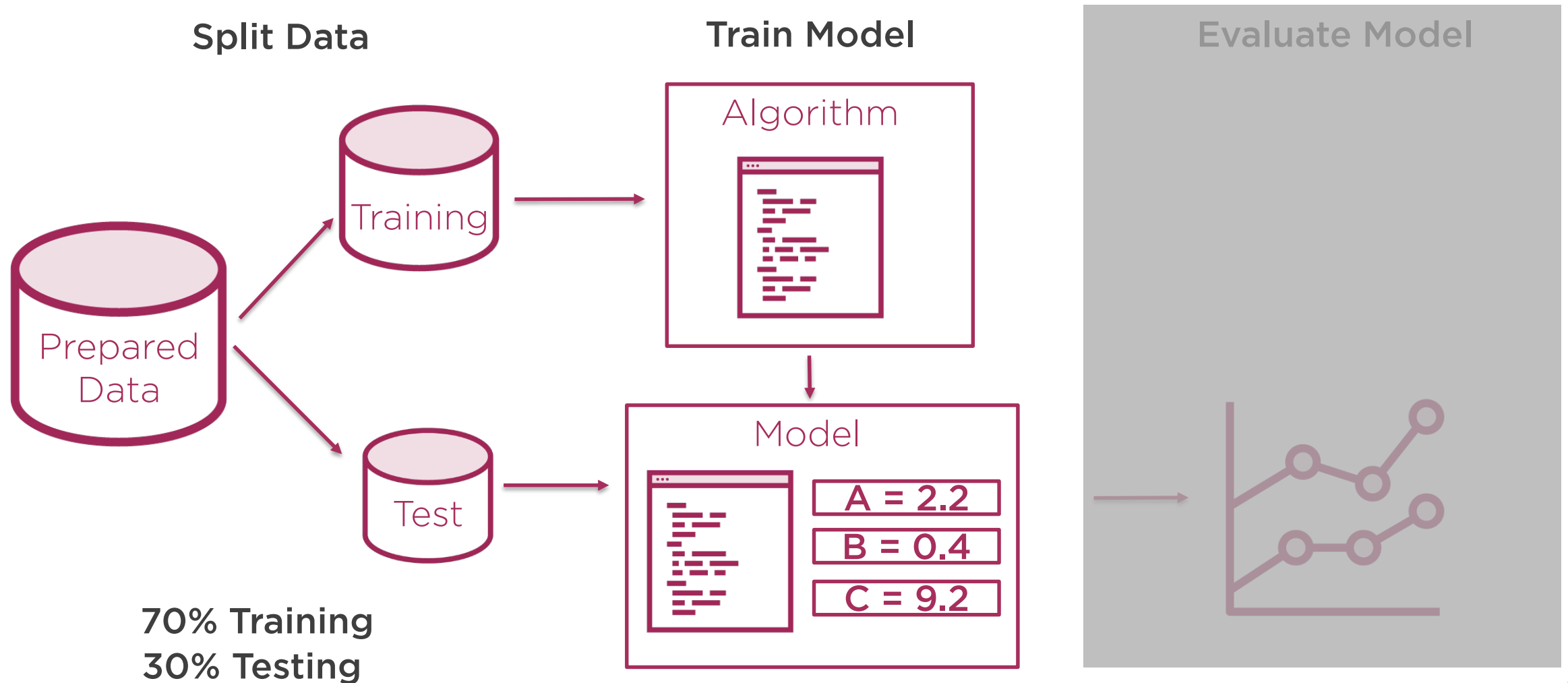**Verify training performance with new data**

Training Overview

Split Data

Train Model

Evaluate Model

Prepared Data

Training

Algorithm

Test

Model

A = 2.2

B = 0.4

C = 9.2

70% Training
30% Testing

What about the test data?

Training Overview

Split Data

Train Model

Real World Model Performance

Training

Prepared Data

Test

Algorithm

Model

A = 2.6

B = 0.5

C = 8.3

70% Training
30% Testing

Training Overview

Selecting Training Features

We want minimum features (columns)

Selected features
- Origin and Destination
- Day of Week
- Carrier
- Departure Time Block
- Arrival Delay 15 (required)

R Training Tip

Don't rewrite from scratch

R package for training available

Caret Package

Caret - Classification And REgression Training

Toolset for training and evaluation tasks
- Data splitting
- Pre-processing
- Feature selection
- Model tuning

Common interface across algorithms

Summary

Reviewed training process

Used R to split data and train model

Utilized the Caret package