

Testing Your Model's Accuracy



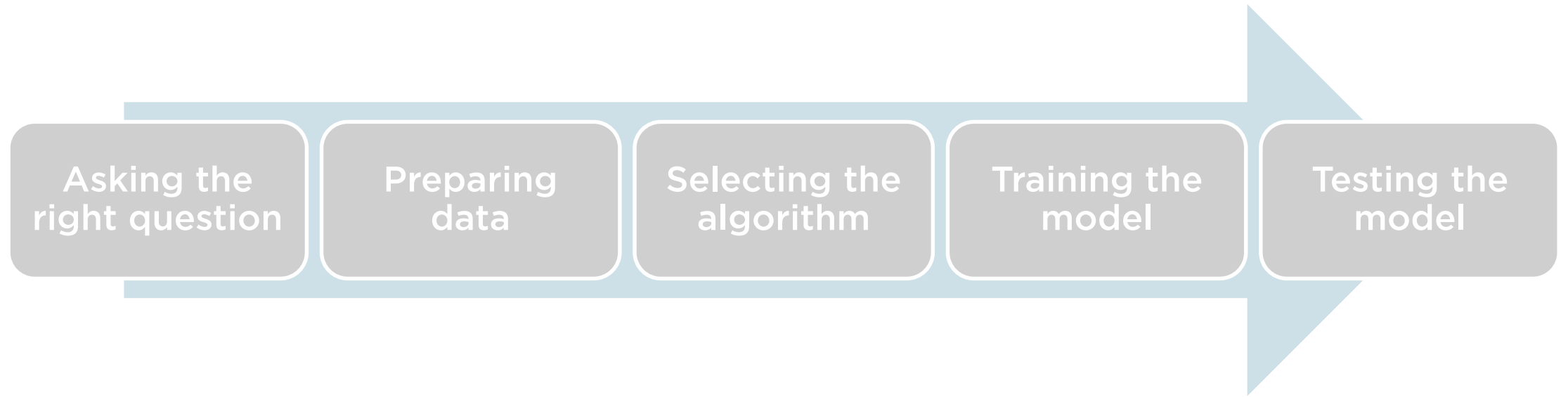
Jerry Kurata

CONSULTANT

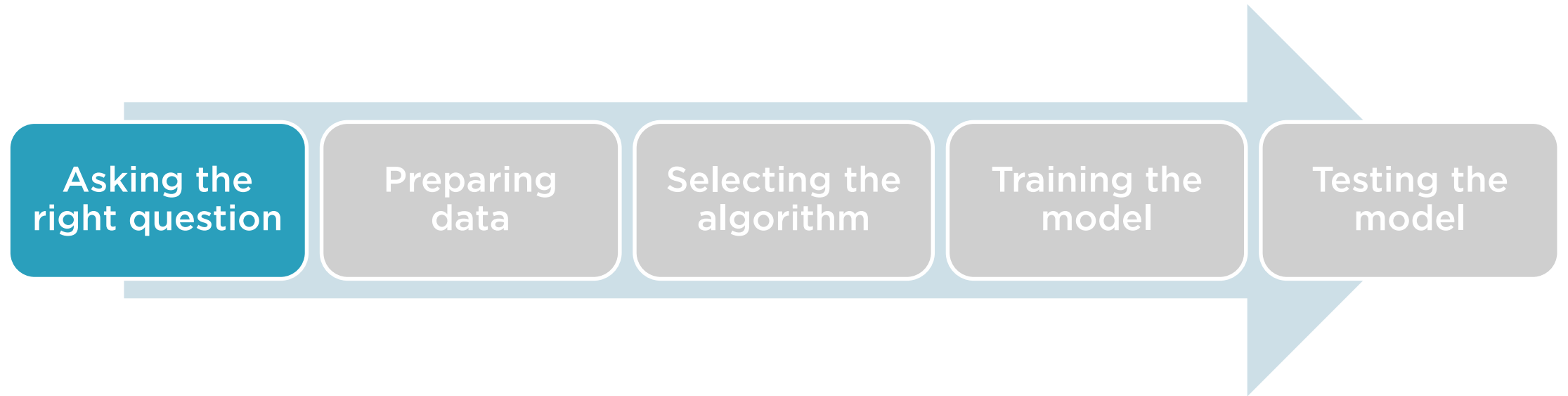
@jerrykur www.insteptech.com



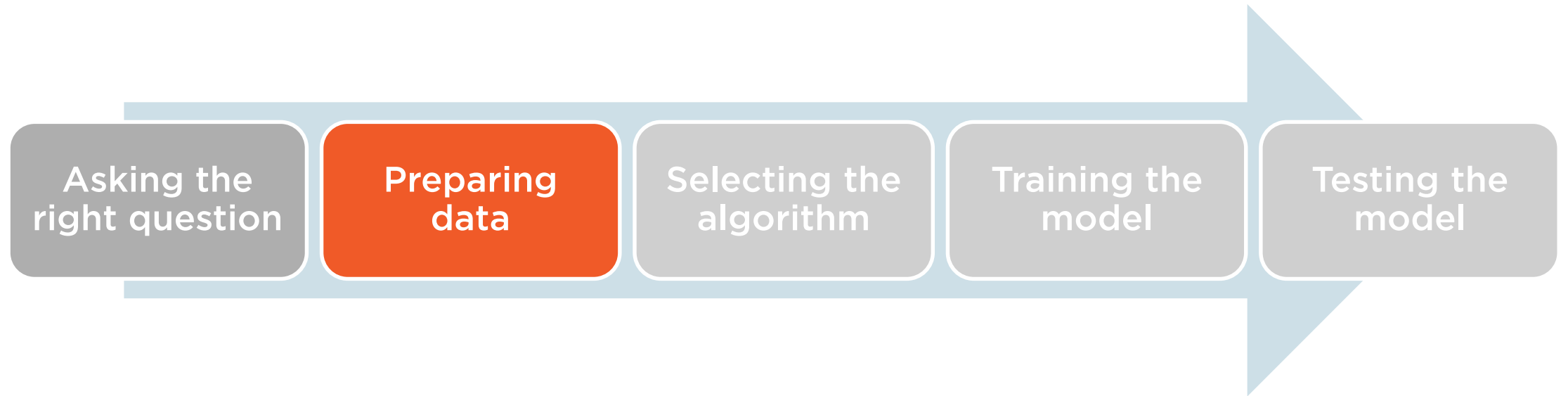
Machine Learning Workflow



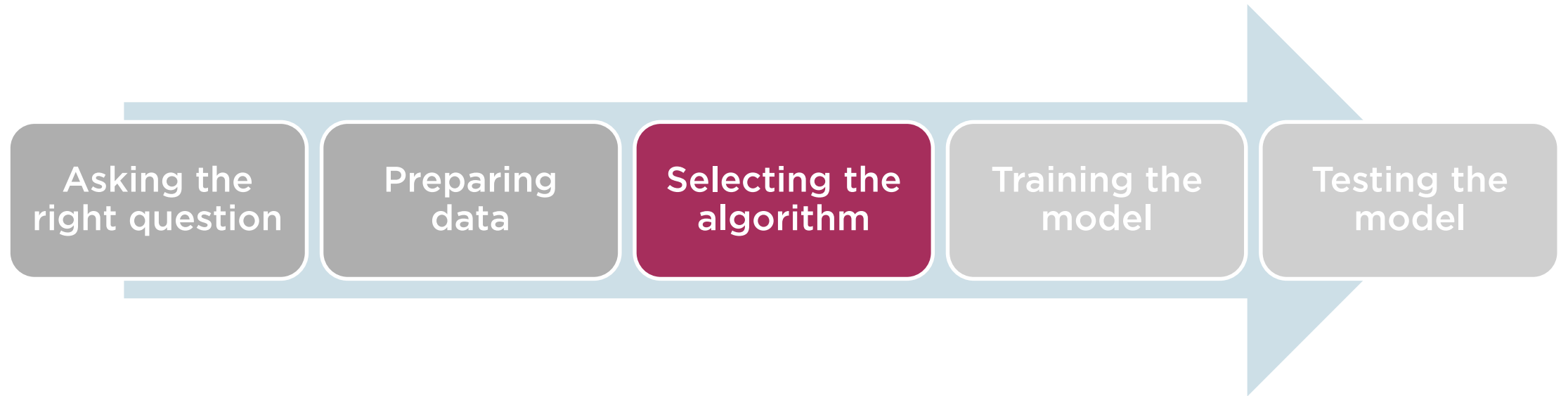
Machine Learning Workflow



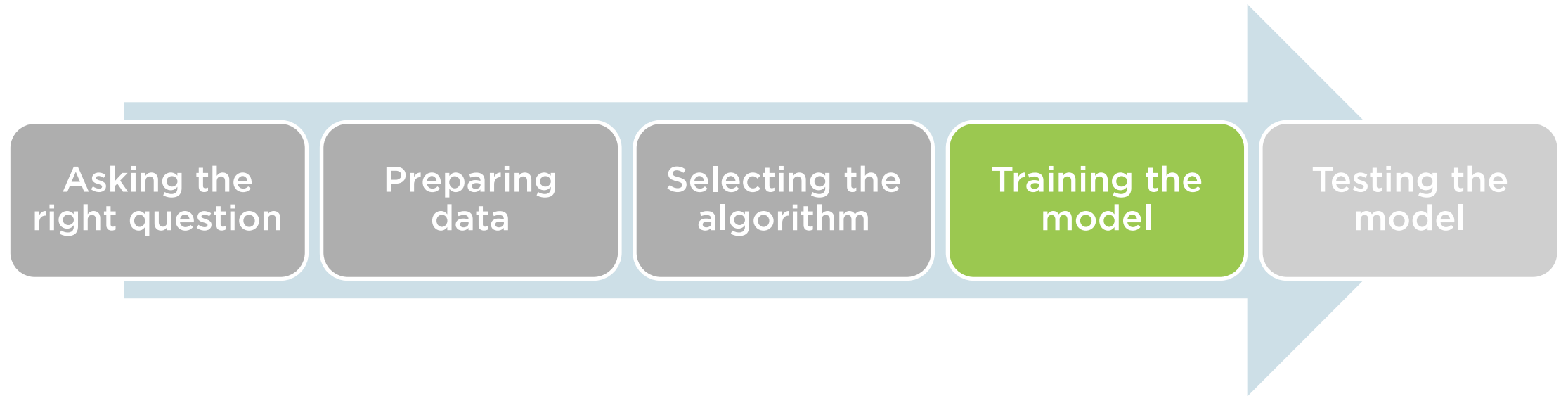
Machine Learning Workflow



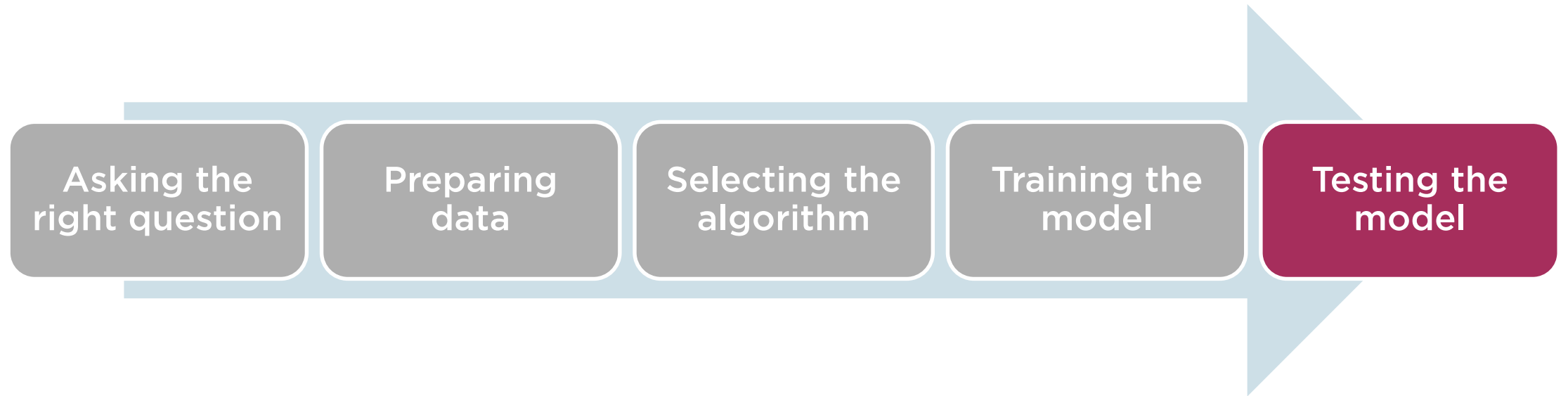
Machine Learning Workflow



Machine Learning Workflow



Machine Learning Workflow



Overview



Evaluate the model against test data

Interpret results

Improve results



Statistic are only data. We
define what is good or bad.





Delayed flights not being predicted

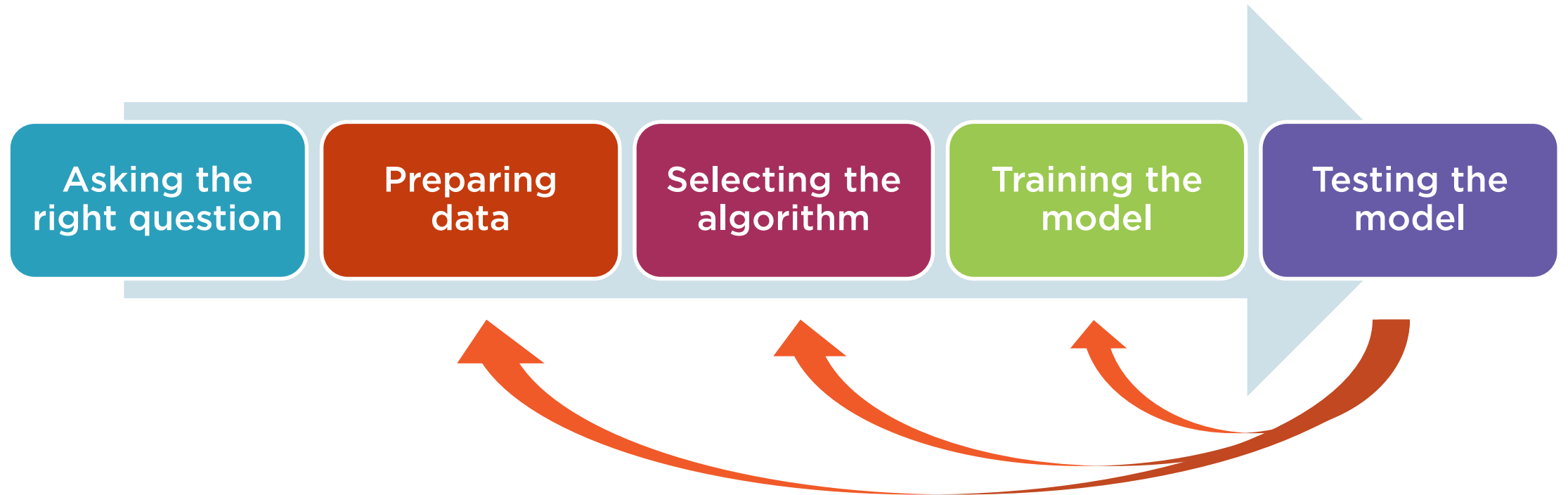
Need to increase prediction accuracy

What are our options?

Follow workflow



Machine Learning Workflow



Options for Improving Performance

Add additional columns

DEP_DEL15

- Will predict arrival delay
- Departure delay detected late
- May not be useful predictor

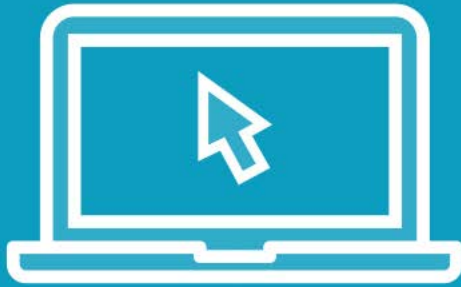
Adjust training settings

Select a better algorithm

Random Forest



Demo



Train and evaluate Random Forest



Options for Improving Performance (Round 2)

Adjust training settings

Select a better algorithm

Rethink the problem

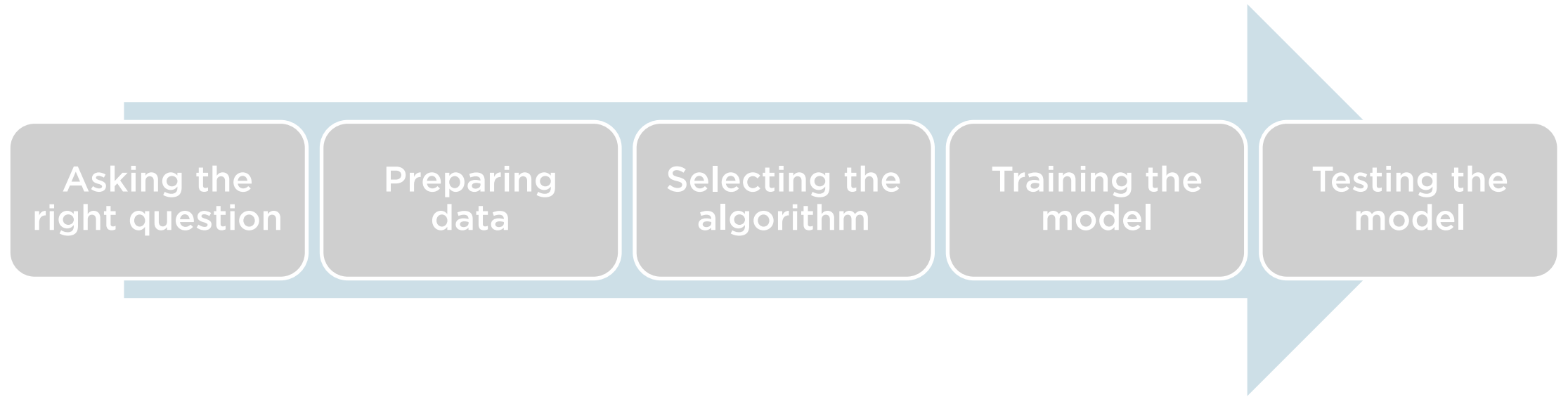
What causes delays?

Subject knowledge required?

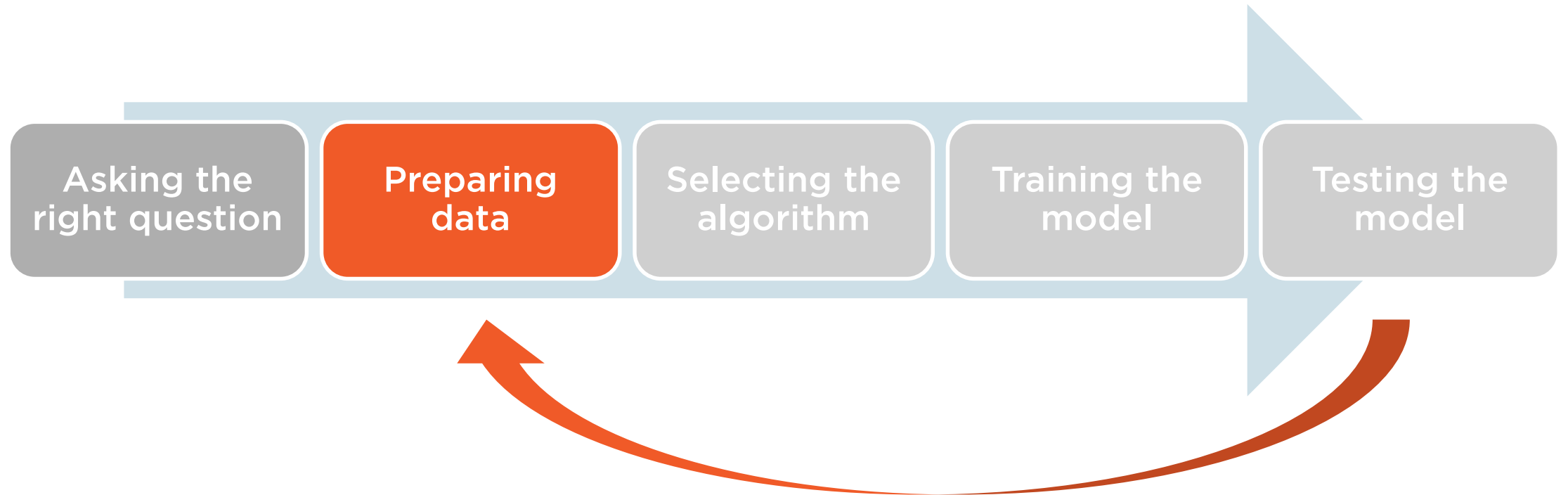
Research shows weather is a delay
factor



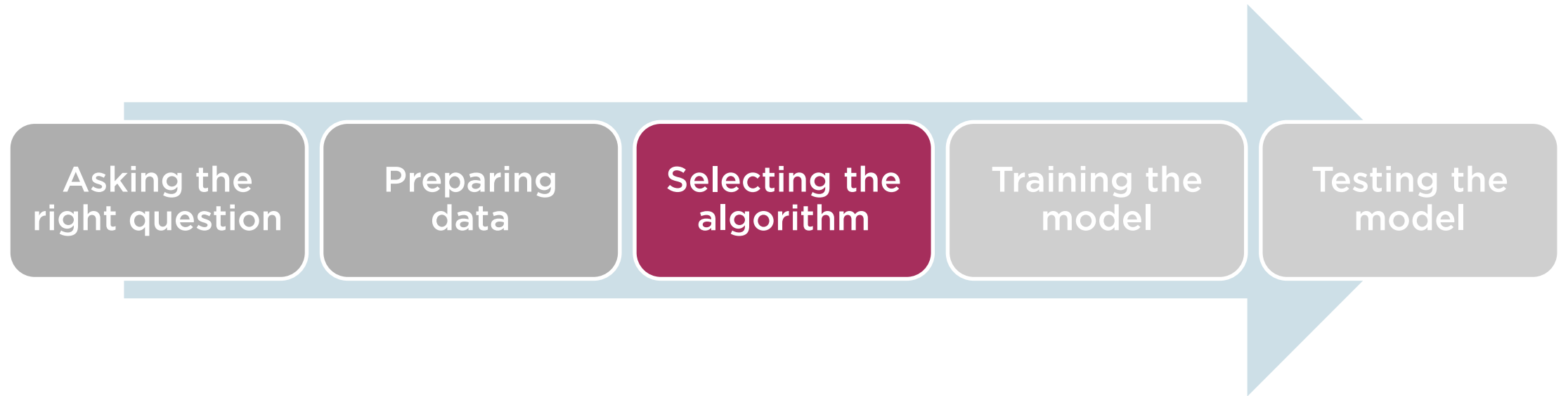
Machine Learning Workflow



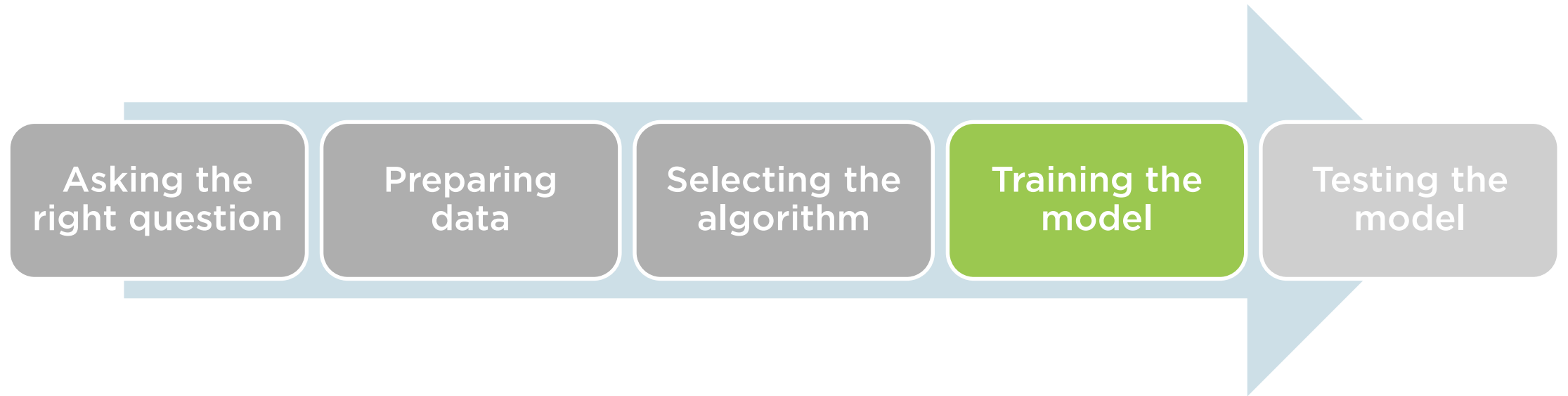
Machine Learning Workflow



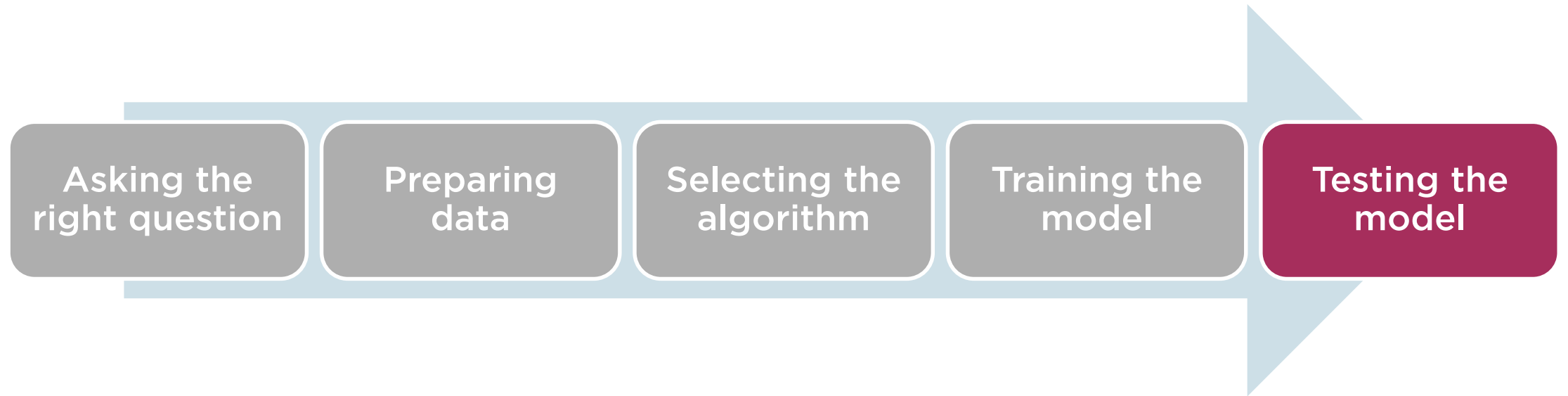
Machine Learning Workflow



Machine Learning Workflow



Machine Learning Workflow



Performance Improvement Cycle

Change data, settings, algorithm or all of
the above

Improve each cycle

The difficult part is knowing when to stop



Summary



Evaluated Logistic Regression model

- `predict()`
- `confusionMatrix()`

Improved performance using Random Forest algorithm

Defined future improvements incorporating weather data