

Asking the Right Question



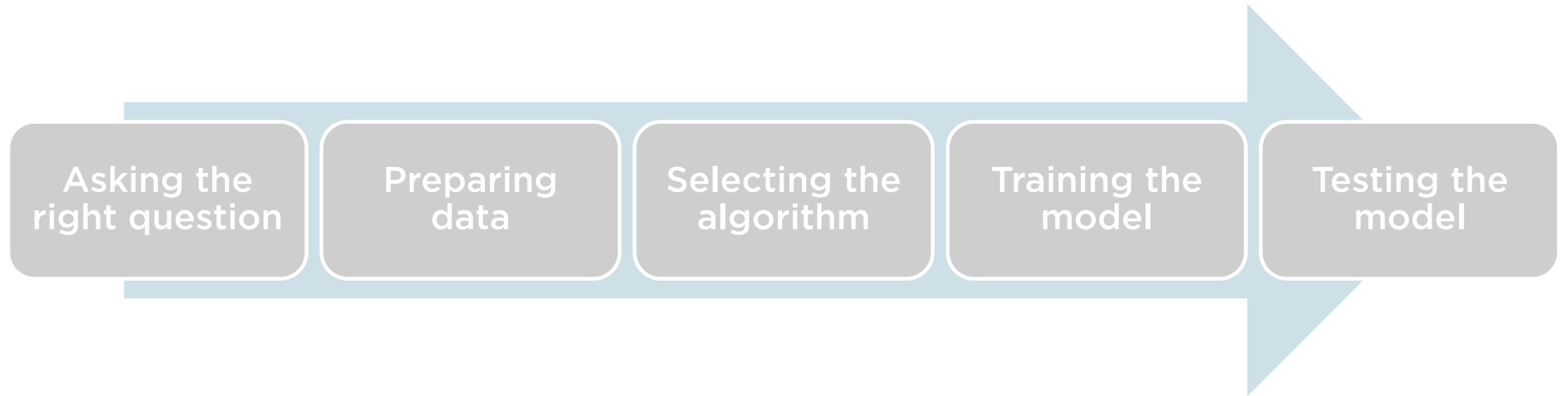
Jerry Kurata

CONSULTANT

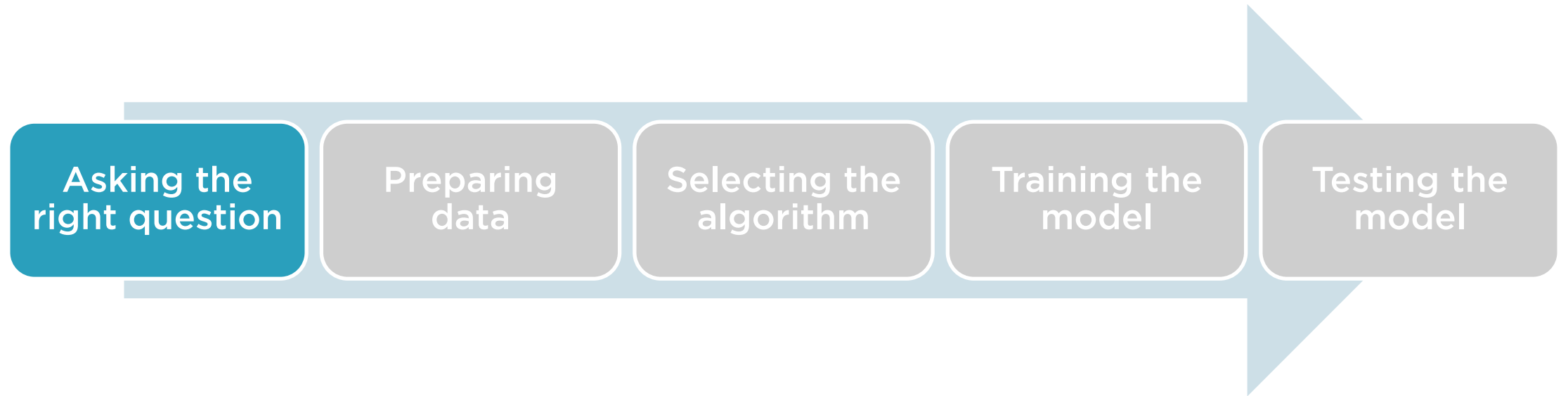
@jerrykur www.insteptech.com



Machine Learning Workflow



Machine Learning Workflow



Don't we
already have
the question?

“Predict if a flight will be on-time”

Need statement to direct and validate work

**Define end goal, starting point, and how to
achieve goal**



Solution Statement Goals

Define scope (including data sources)

Define target performance

Define context for usage

Define how solution will be created



Scope and Data Sources

“Predict if a flight would be on-time”

“Using DOT data, predict if a flight would be on-time”

US flights only

Flights between US airports only

DOT database is a good source



Data

“Using DOT data, predict if a flight would be on-time”

“Using DOT data, predict if a flight would be delayed”

Preliminary data review

Delays tracked, not on-time



Performance Targets

“Using DOT data, predict if a flight would be delayed”

“Using DOT data, predict with 70+% accuracy if a flight would be delayed”

Binary result (True or False)

70% Accuracy is common target



Context

“Using DOT data, predict with 70+% accuracy if a flight would be delayed”

“Using DOT data, predict with 70+% accuracy if a flight would arrive 15+ minutes after the scheduled arrival time.”

Data driven results

DOT “delayed” => greater than 15 minutes after scheduled



Solution Creation

“Using DOT data, predict with 70+% accuracy if a flight would arrive 15+ minutes after the scheduled arrival time.”

“Use the Machine Learning Workflow to process and transform DOT data to create a prediction model. This model must predict whether a flight would arrive 15+ minutes after the scheduled arrival time with 70+% accuracy.”

Machine Learning Workflow

- Process DOT data
- Transform data as required



Use the Machine Learning Workflow to process and transform DOT data to create a prediction model. This model must predict whether a flight would arrive 15+ minutes after the scheduled arrival time with 70+% accuracy.



Use the Machine Learning Workflow to process and transform DOT data to create a prediction model. This model must predict whether a flight would arrive 15+ minutes after the scheduled arrival time with 70+% accuracy.



Use the Machine Learning Workflow to process and transform DOT data to create a prediction model. This model must predict whether a flight would arrive 15+ minutes after the scheduled arrival time with 70+% accuracy.



Use the Machine Learning Workflow to process and transform DOT data to create a prediction model. This model must predict whether a flight would arrive 15+ minutes after the scheduled arrival time with 70+% accuracy.



Use the *Machine Learning Workflow* to process and transform DOT data to create a *prediction model*. This model must predict whether a flight would arrive 15+ minutes after the scheduled arrival time with 70+% accuracy.

