

Building Your First Machine Learning Solution

GETTING OUR FEET READY TO RUN



Mohammed Osman

SENIOR SOFTWARE DEVELOPER

@cognitiveosman www.cognitiveosman.com



Overview



Machine Learning: What and why?

What is different about Machine Learning

Types of Machine Learning algorithms

Machine Learning workflow

Introducing real-world problem

Environment setup



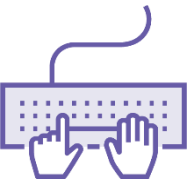
What to Expect?



No **Machine Learning** or **Python background** is required



Important Machine Learning concepts



Hands-on based



Machine Learning

Machine learning is the **scientific study** of **algorithms and statistical models** that computer systems use to perform **a specific task** without using **explicit instructions**, relying on **patterns and inference** instead. (Wikipedia)



Machine Learning under Lens

Scientific Study

Algorithms and
Statistics

Specific Task

No Explicit
Instructions

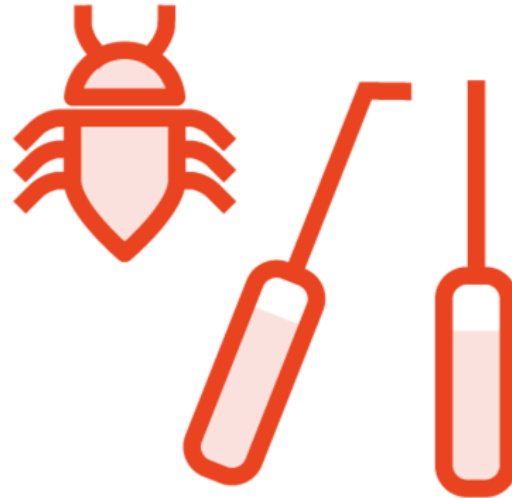
Patterns and
Inferences



Some Applications of Machine Learning



Image Recognition



Malware Detection



Voice Recognition



Why Machine Learning Is so Hot?



High
Dimensionality
Challenge



Data Flood



Computation
Power Growth



Matured Field

Data is the new oil



Traditional Programming vs. Machine Learning

Traditional Programming

We have **Full Understanding** of the domain

Solution rules are **static**

Does not require historical data

Straight forward and obvious

Machine Learning

We have **Vague Understanding** of the domain

Solution rules are **dynamic**

Does require historical data

Vague and **tricky** to understand



Machine Learning Algorithms Types

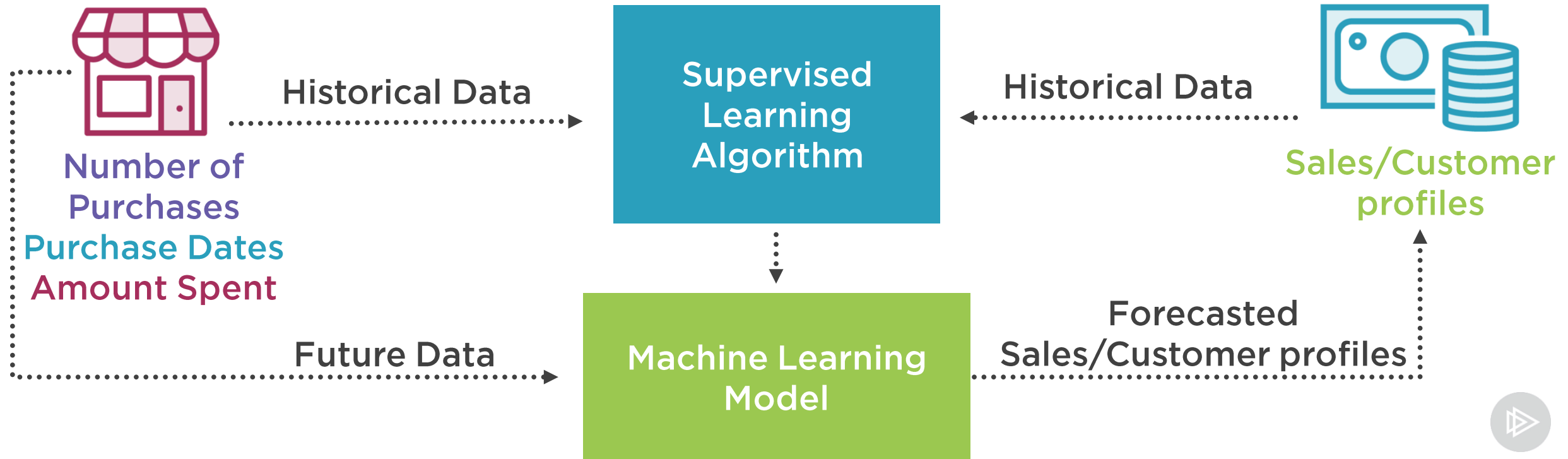
**Supervised
Learning**

**Unsupervised
Learning**

**Reinforcement
Learning**



Supervised Learning



Supervised Learning Types



Regression

Sales forecast: 13,666 \$

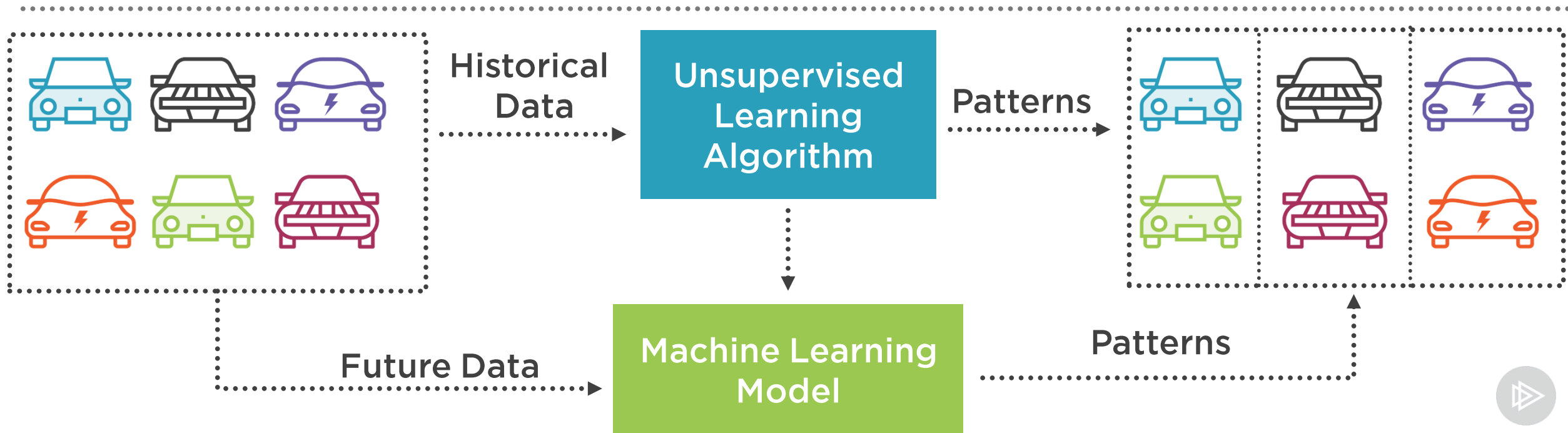


Classification

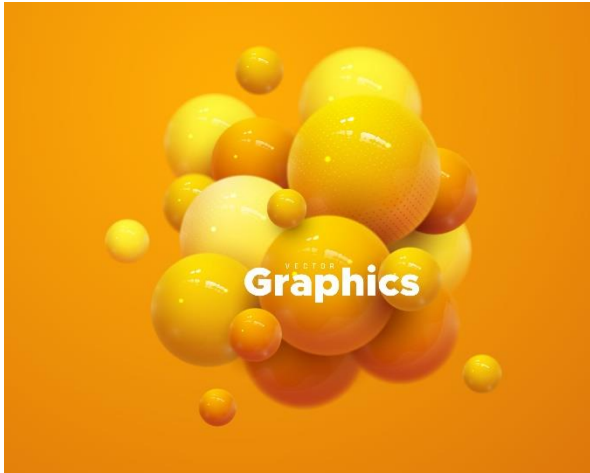
Customers classifications: High
Profile, Medium and Normal



Unsupervised Learning



Unsupervised Learning Types



Clustering



Association



Reinforcement Learning



Learning Types in a Nutshell

	Supervised	Unsupervised	Reinforcement
Objective	Predict future values or categories	Organize data based on underlying structure	Adapt based on the rewards and state from the surrounding environment
Learning source	Output dataset	Input data patterns	Environment state and rewards



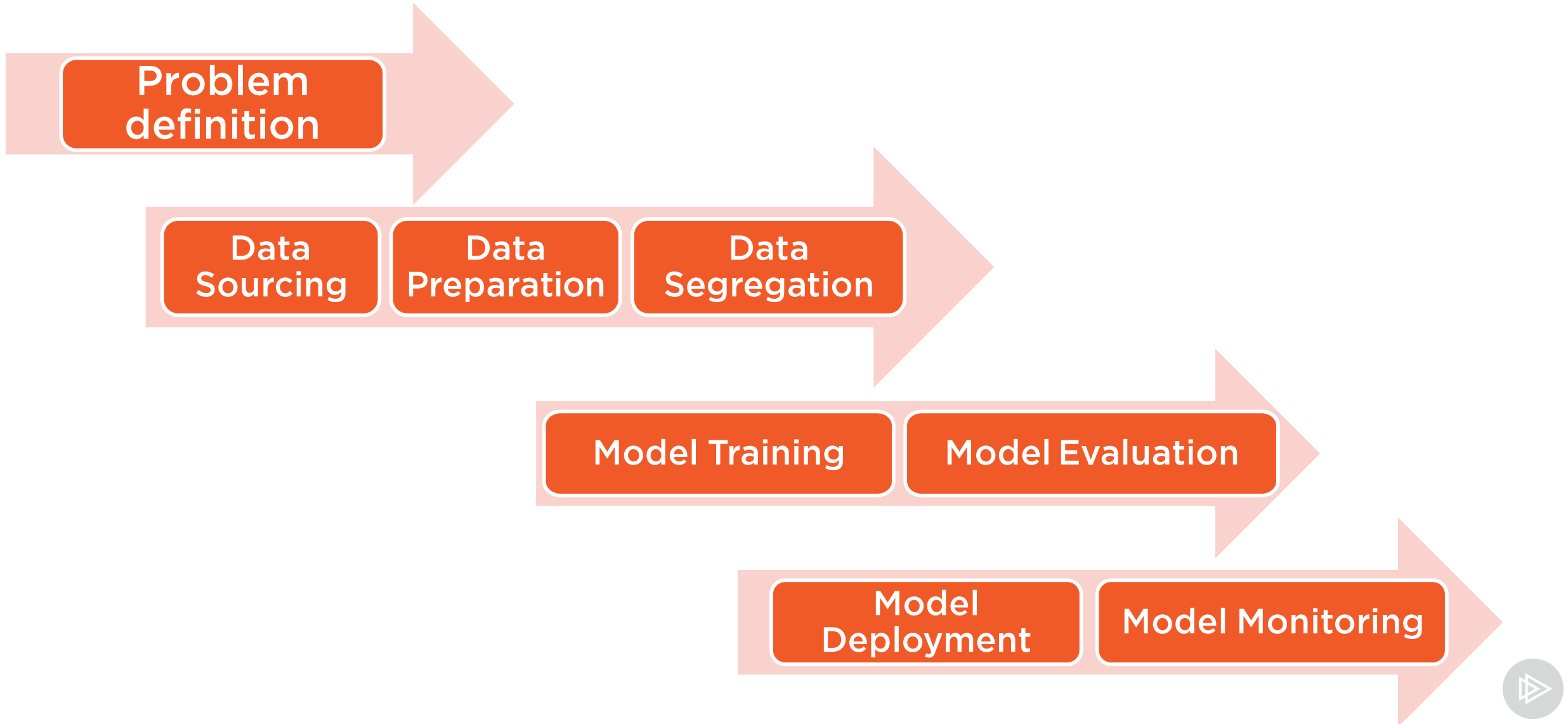
Learning Modes

Batch Learning

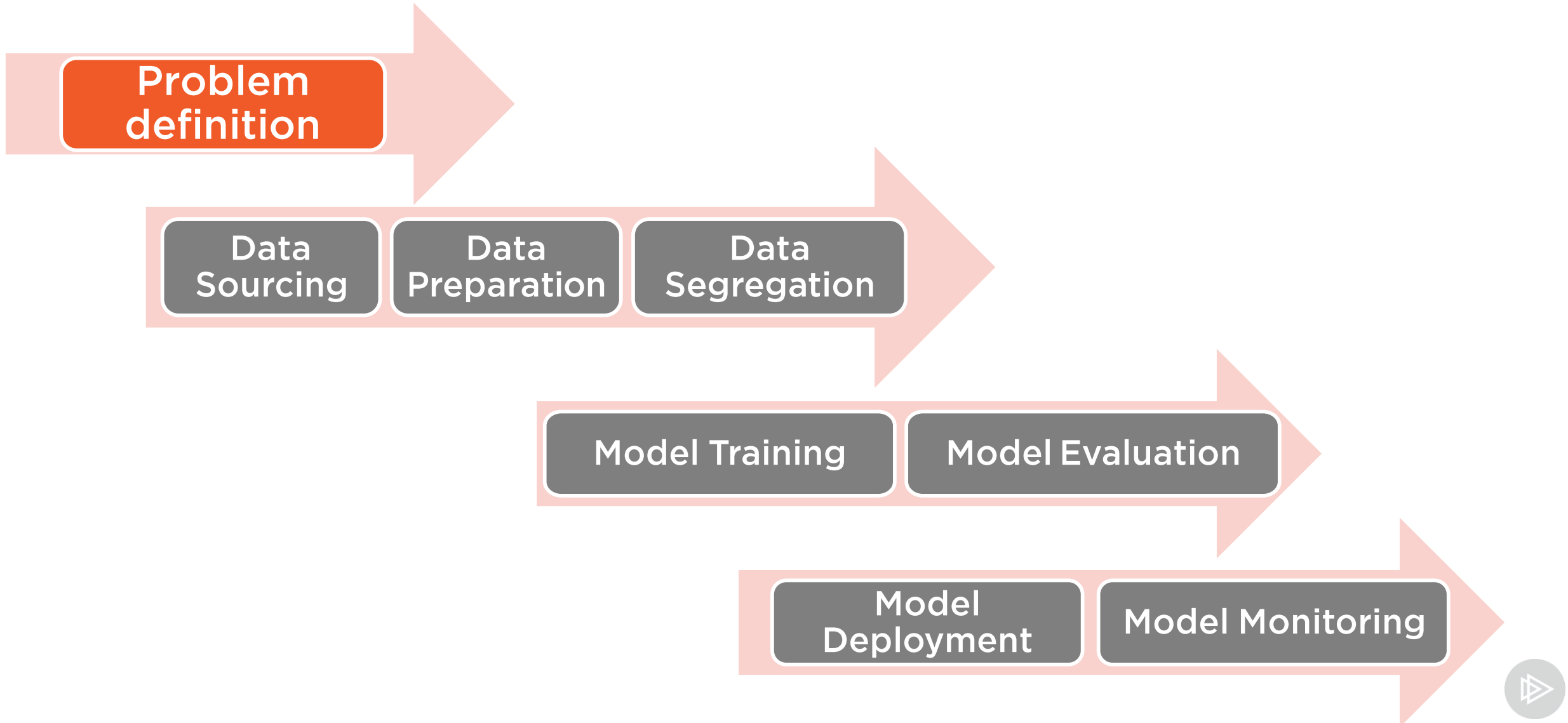
Online Learning



Machine Learning Pipeline



Problem Definition



Problem Definition

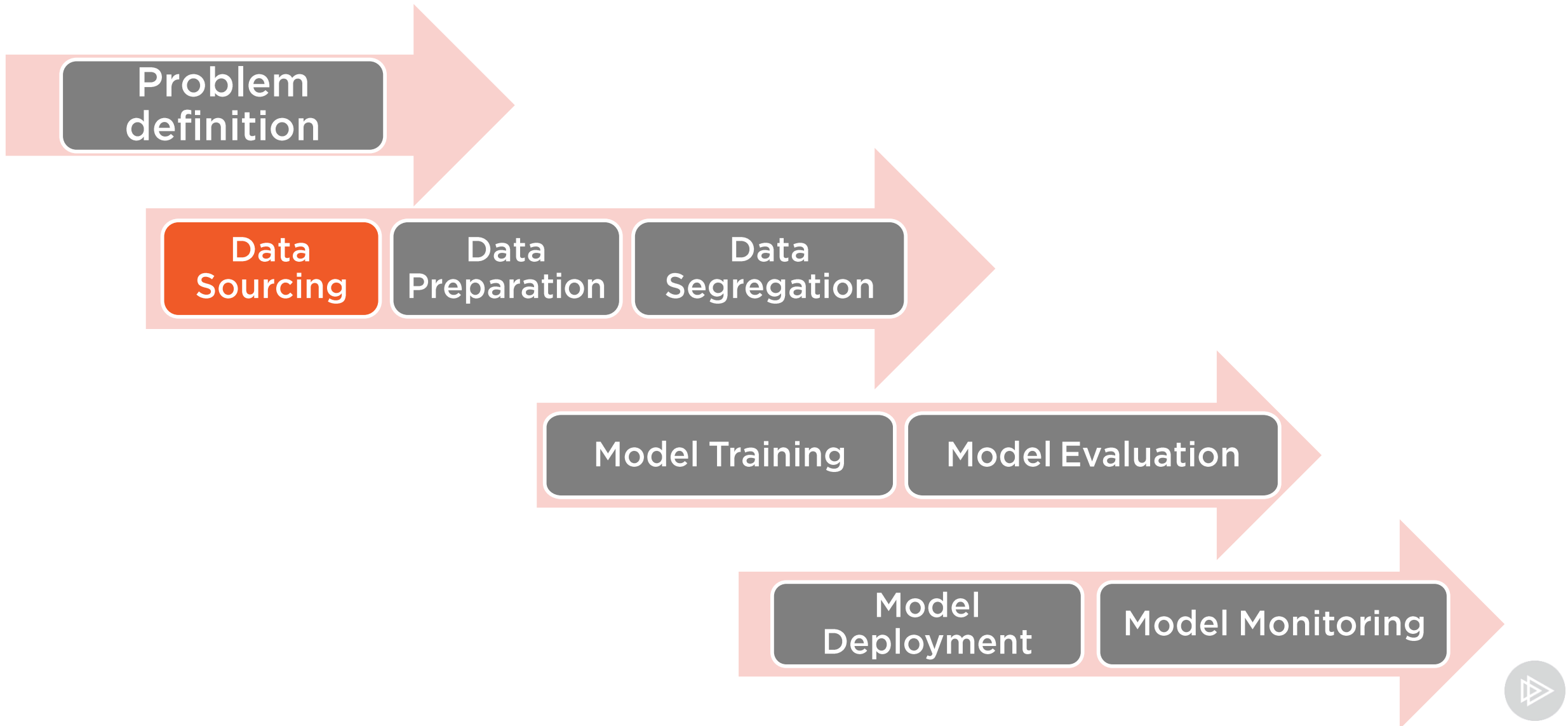


Five questions ML can answer

- Is this A or B? Will this customer buy or not?
- Is this weird?
- How much – or – How many? How many items will I see in the upcoming quarter?
- How is this organized? What are the different customer categories do I have?
- What should I do next?



Data Sourcing



Data Sourcing

{JSON}

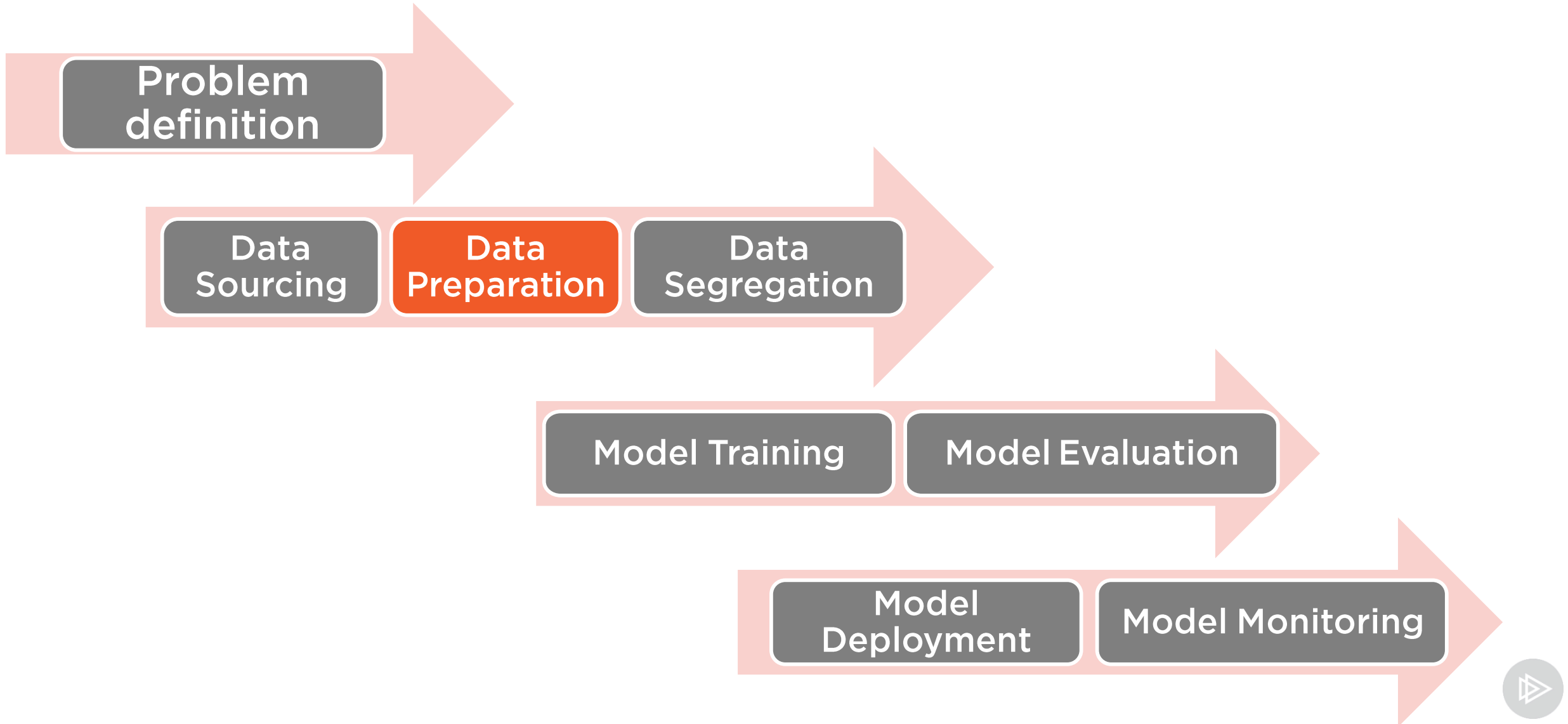


Several data sources may exist

- RESTful Endpoints
- File integration
- SOAP Endpoints
- SQL Table
- Sensors with proprietary format



Data Preparation



Data Preparation

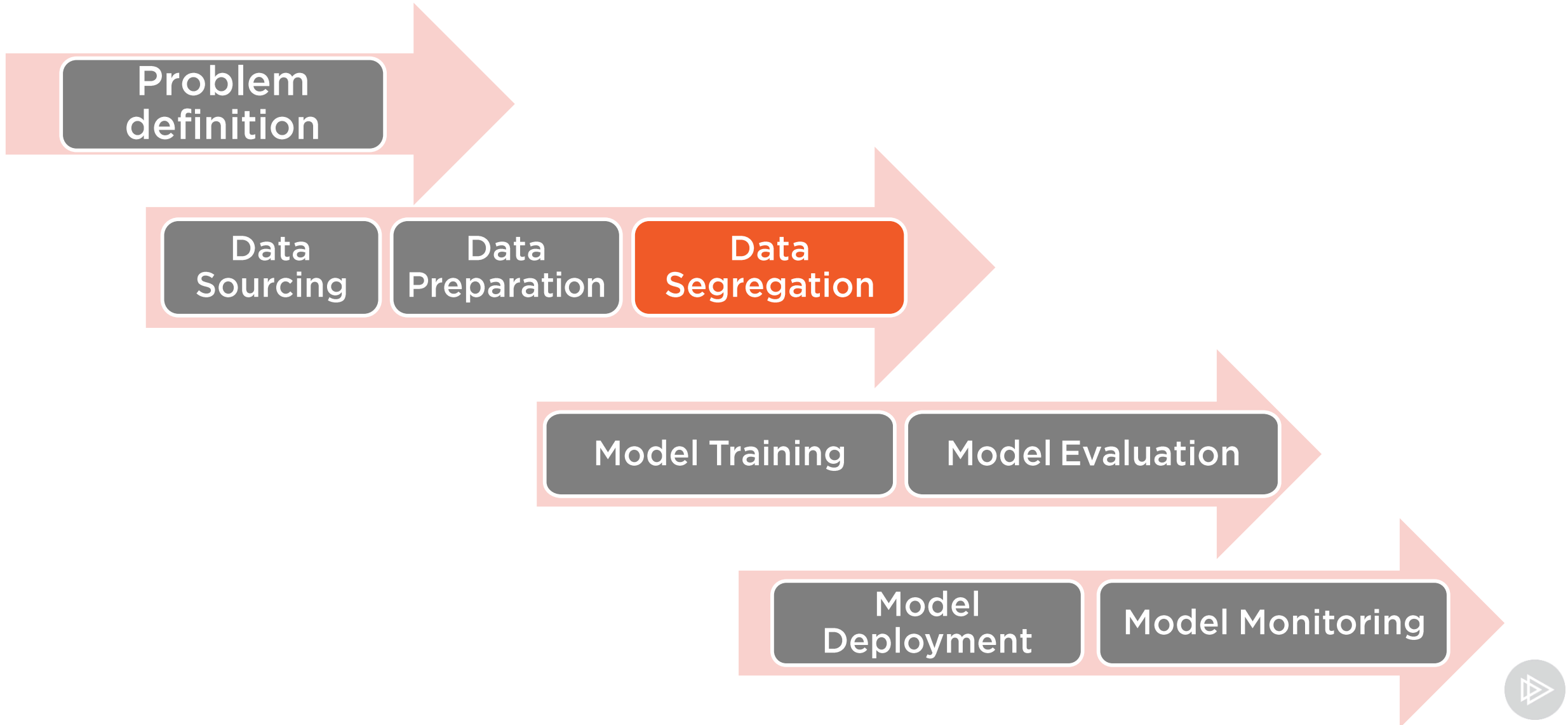


Several data preparation actions

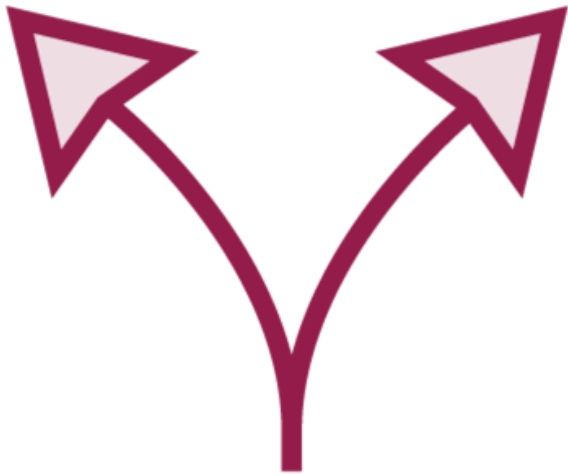
- Dealing with missing data
- Dropping unnecessary attributes
- Detecting outliers
- Etc.



Data Segregation



Data Segregation

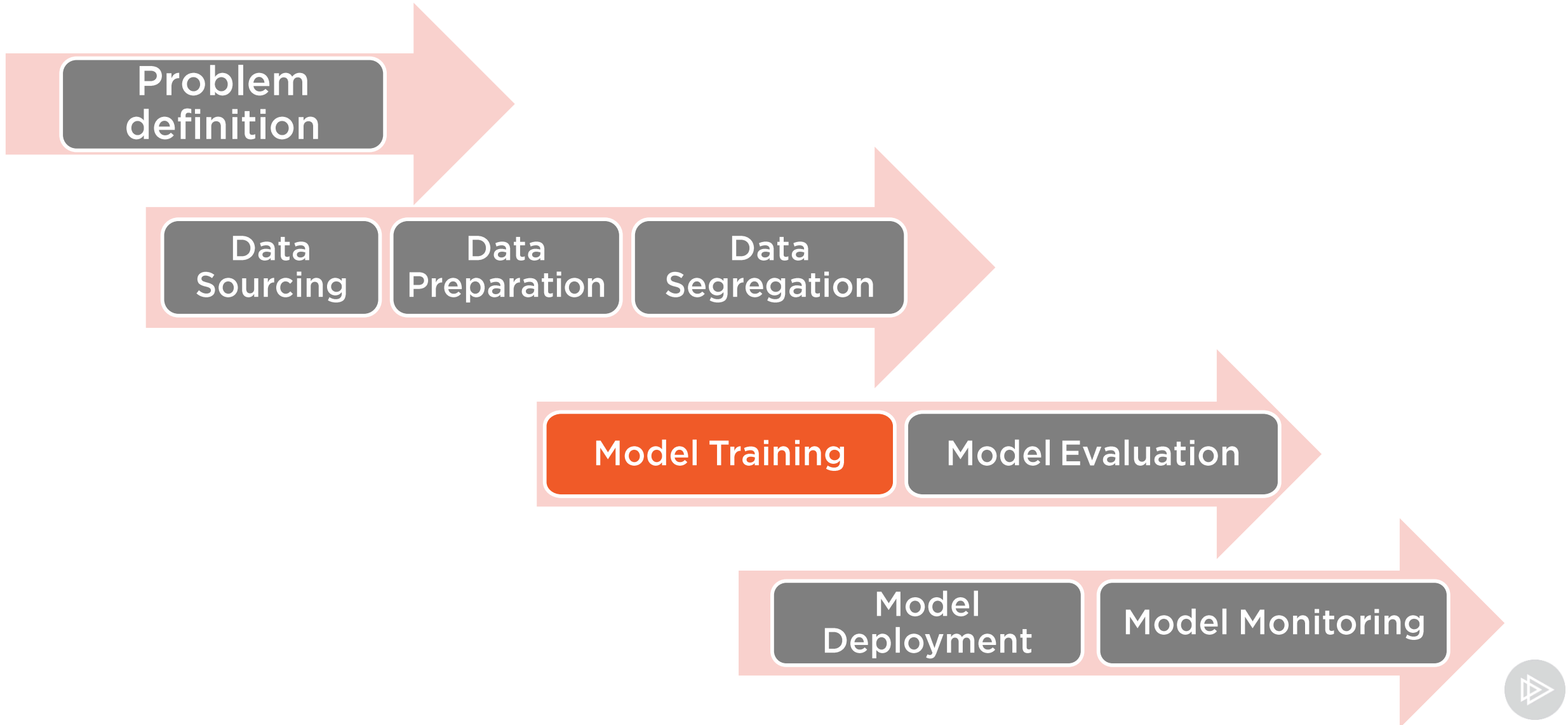


Data is segregated to

- Training set
- Validation set



Model Training



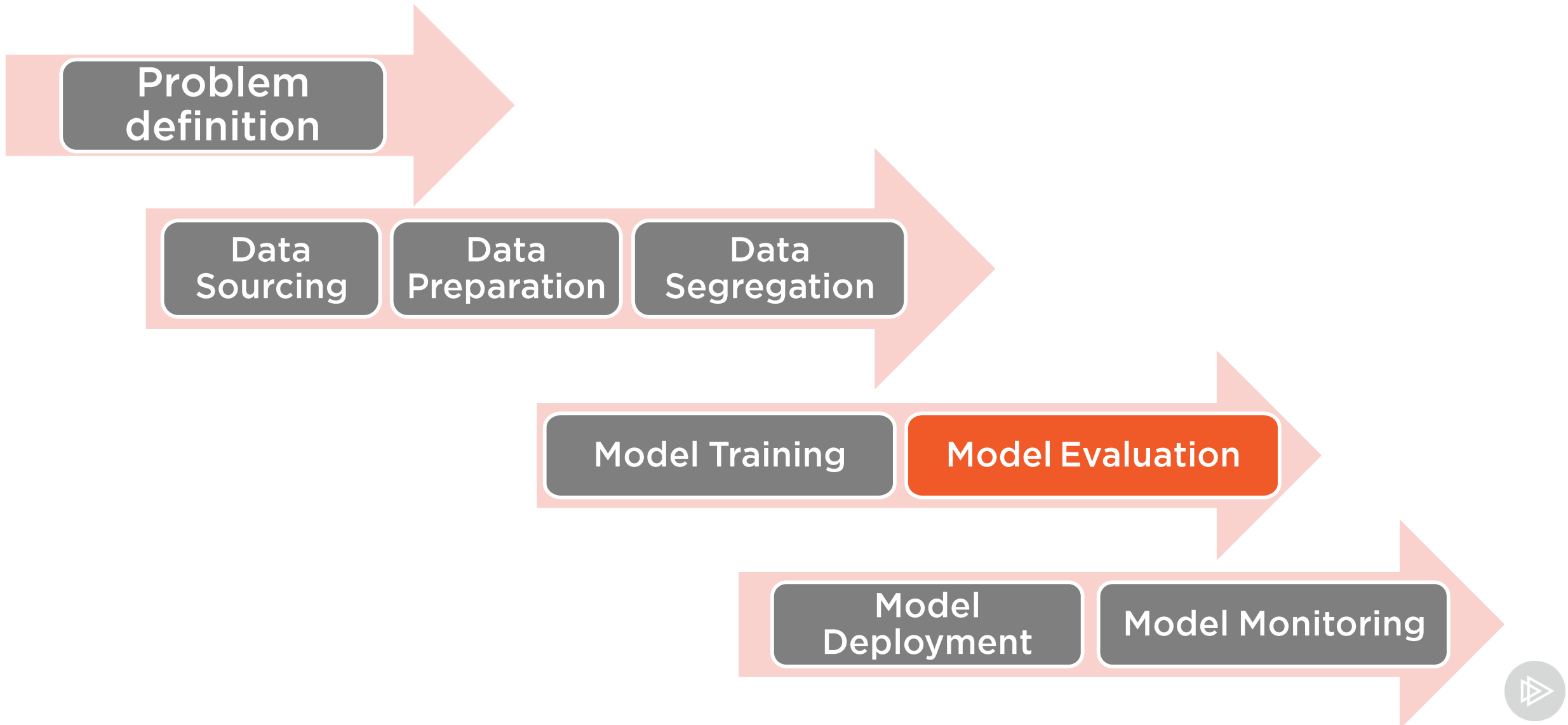
Model Training



Machine Learning algorithm adjusts
Usually on-shelf recipe
May need trying several algorithms



Model Evaluation



Model Evaluation

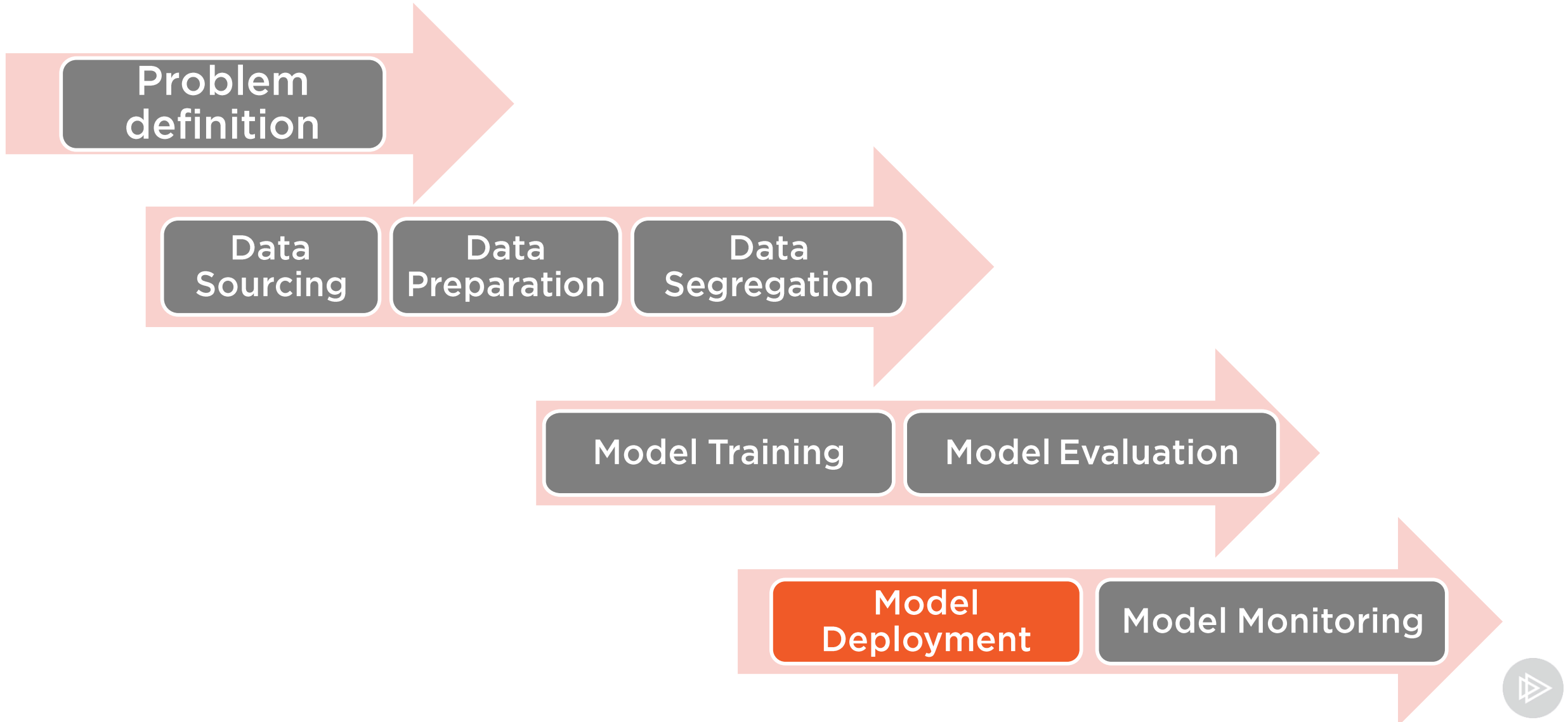


Examining model performance using validation data

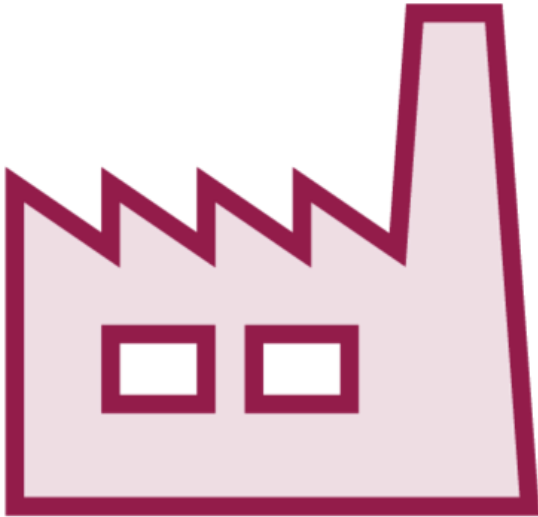
Different performance measures based on the algorithm type



Model Deployment



Model Deployment

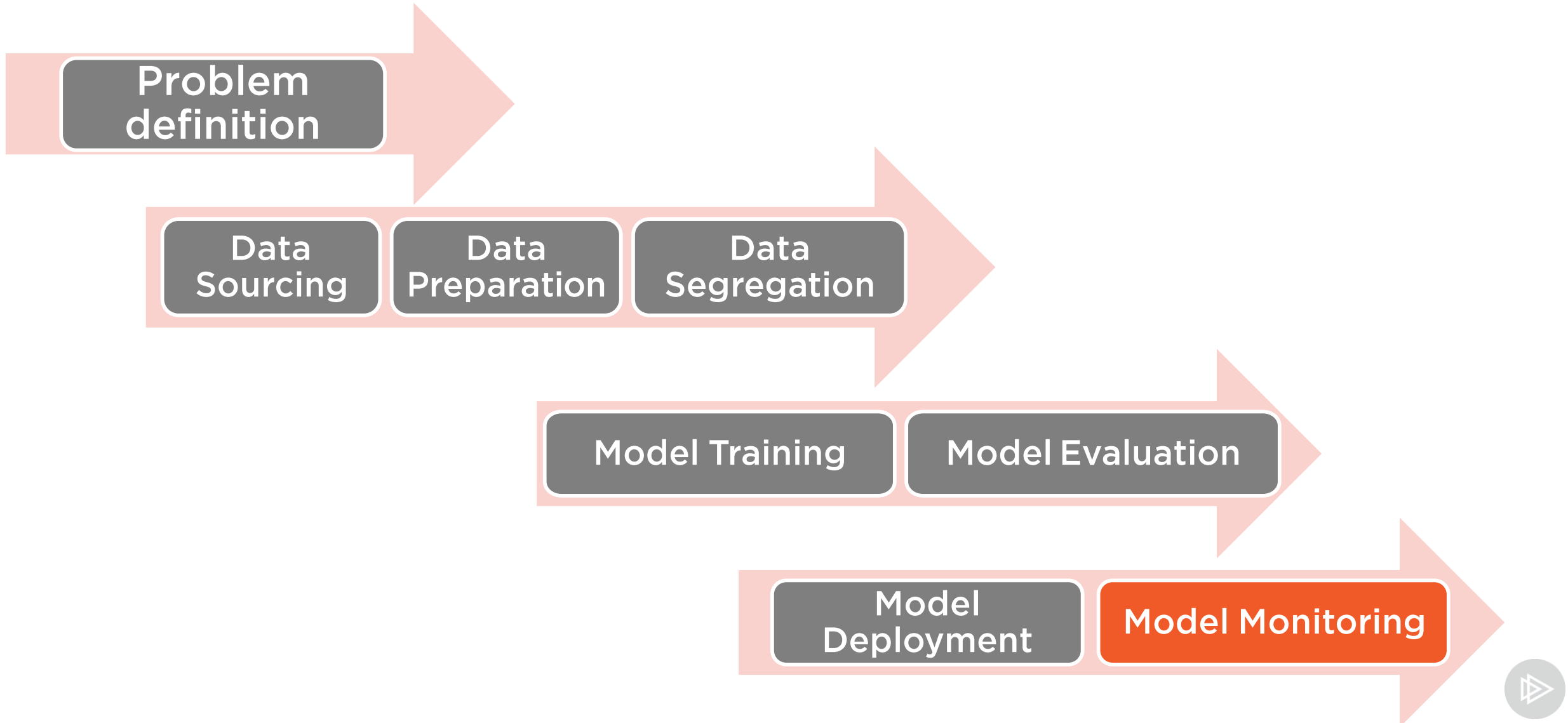


Making model useful for business

Deployed Machine Learning model can take several formats



Model Monitoring



Model Monitoring



Making sure our model continues to do good



Problem Definition: Forest Fire Forecast

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind	rain	area
6	5	mar	mon	90.1	39.7	86.6	6.2	15.2	27	3.1	0	31.86
8	6	aug	sun	90.2	99.6	631.2	6.3	16.2	59	3.1	0	32.07
3	4	sep	fri	93.3	141.2	713.9	13.9	18.6	49	3.6	0	35.88
4	3	mar	mon	87.6	52.2	103.8	5	11	46	5.8	0	36.85
2	2	jul	fri	88.3	150.3	309.9	6.8	13.4	79	3.6	0	37.02
7	4	sep	wed	90.1	82.9	735.7	6.2	15.4	57	4.5	0	37.71
4	4	sep	sun	93.5	149.3	728.6	8.1	22.9	39	4.9	0	48.55
7	5	oct	mon	91.7	48.5	696.1	11.1	16.1	44	4	0	49.37
8	6	aug	sat	92.2	81.8	480.8	11.9	20.1	34	4.5	0	58.3
4	6	sep	sun	93.5	149.3	728.6	8.1	28.3	26	3.1	0	64.1
8	6	aug	sat	92.2	81.8	480.8	11.9	16.4	43	4	0	71.3
4	4	sep	wed	92.9	133.3	699.6	9.2	26.4	21	4.5	0	88.49
1	5	sep	sun	93.5	149.3	728.6	8.1	27.8	27	3.1	0	95.18
6	4	sep	tue	91	129.5	692.6	7	18.7	43	2.7	0	103.39
9	4	sep	tue	84.4	73.4	671.9	3.2	24.3	36	3.1	0	105.66



We will solve the forest fire
forecast problem by using
the supervised learning, in
particular regression



Demo



Installing Python



Summary



Machine Learning is everywhere

How Machine Learning differs from regular programming

Different types of Machine Learning algorithms

Machine Learning pipeline

Preparing to applying hands-on using Python

