



Machine Learning Models



Scott Pletcher

INSTRUCTOR

What is a Model?

Taking a **problem** or challenge as described by **lots of data**, adding a machine learning **algorithm** and through **computation**, trying to figure out a mathematical **formula** that can **accurately generalize** about that problem.



What is a Model?

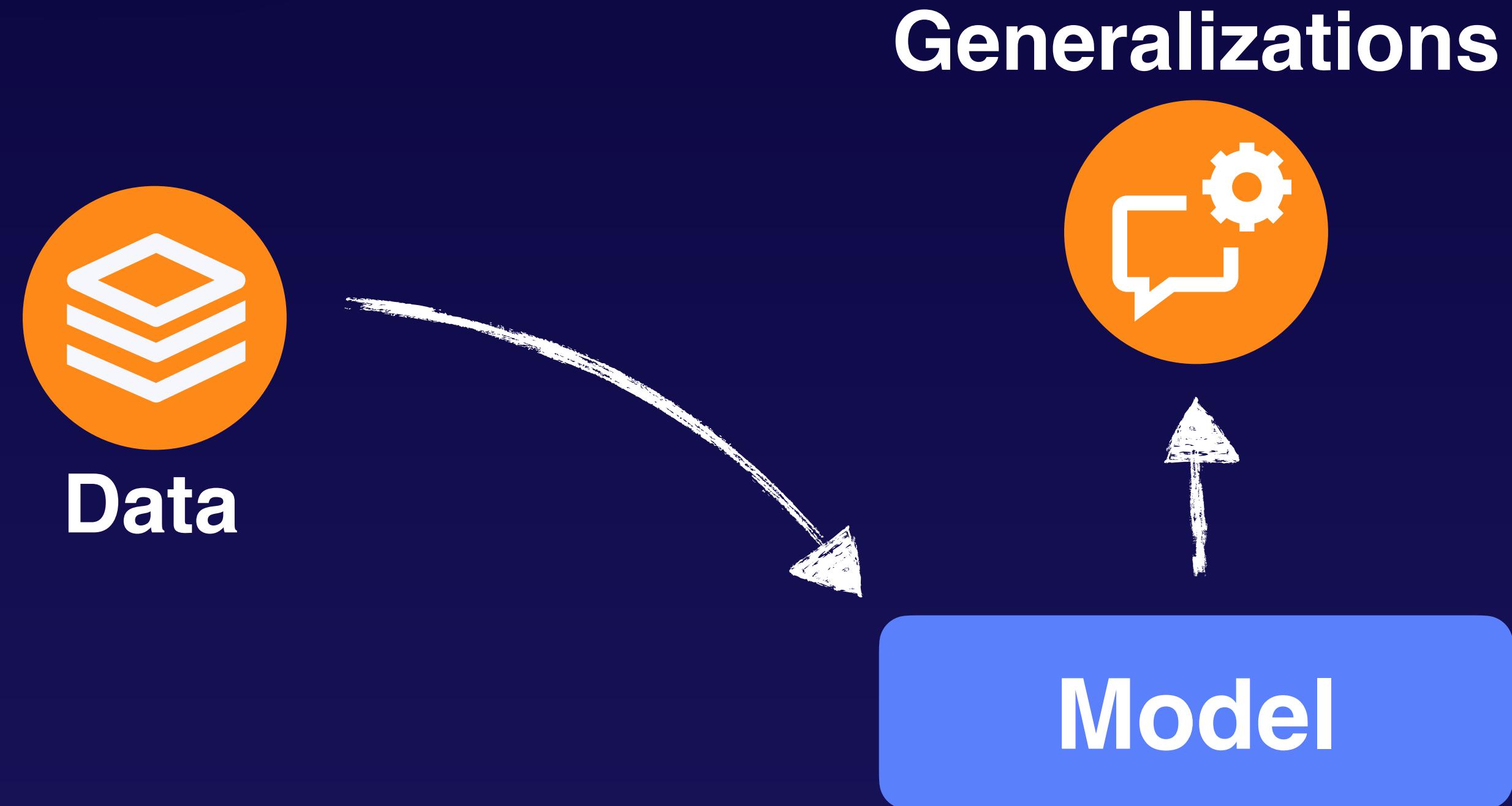
Model

Generalizations

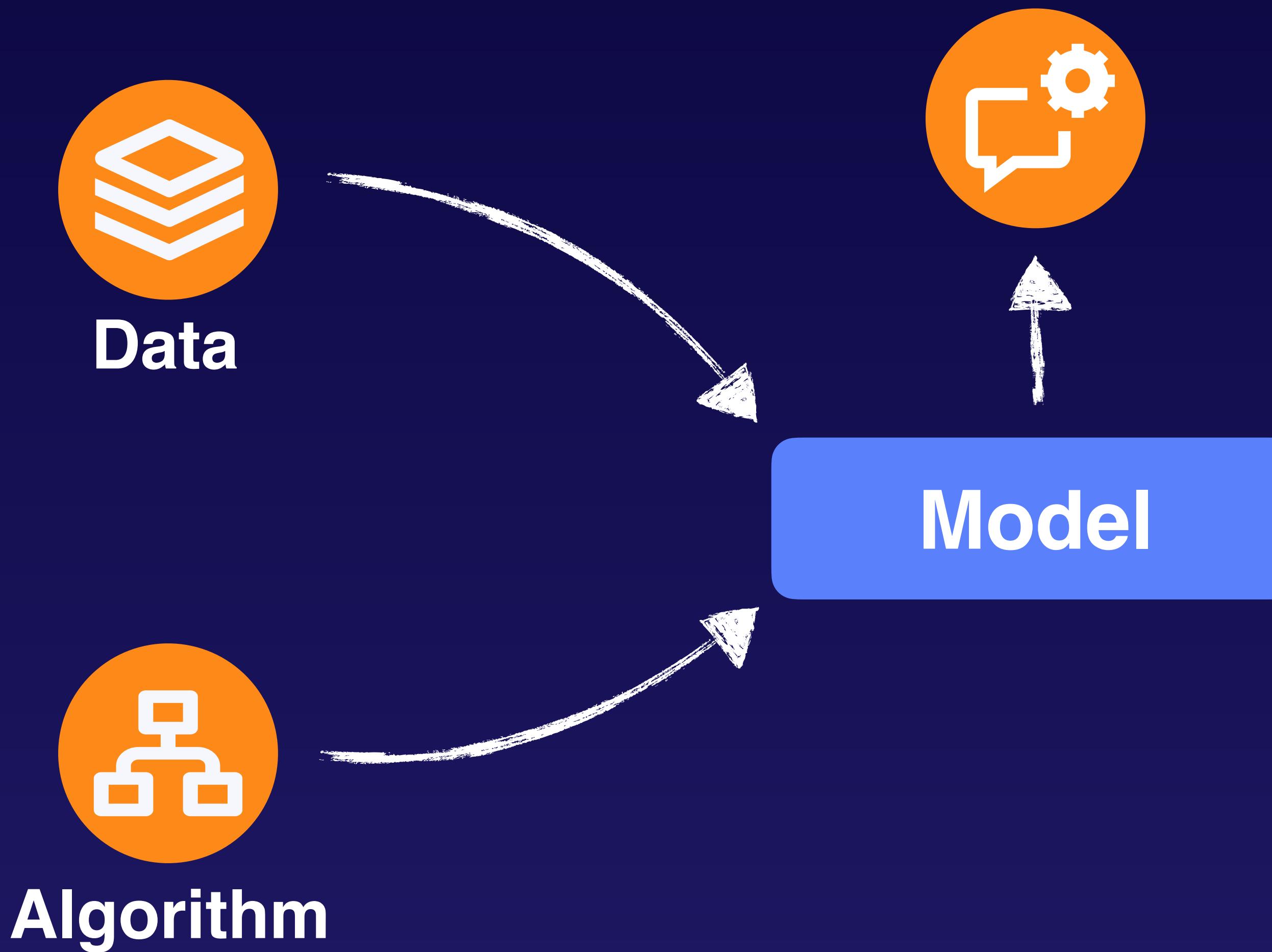


Model

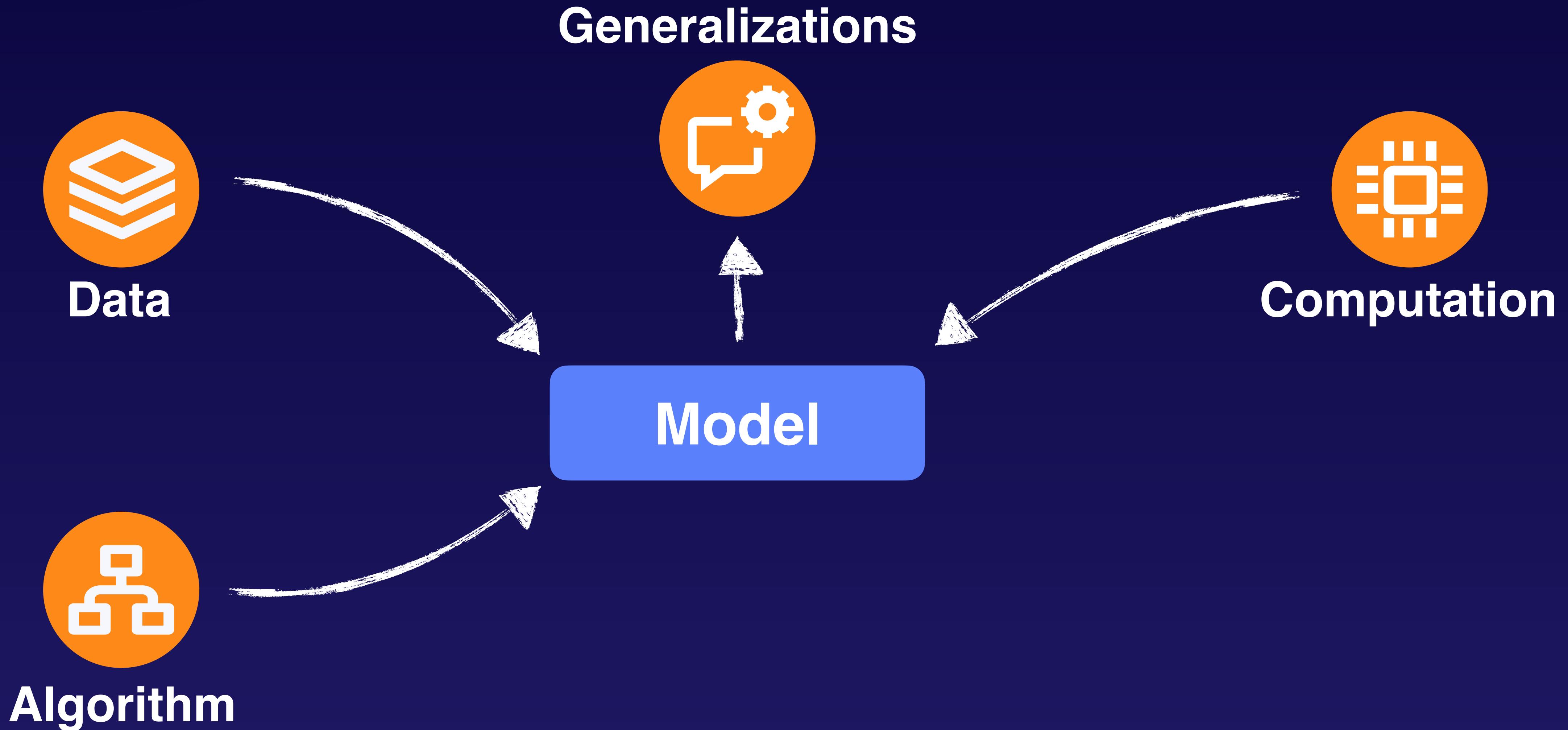
What is a Model?



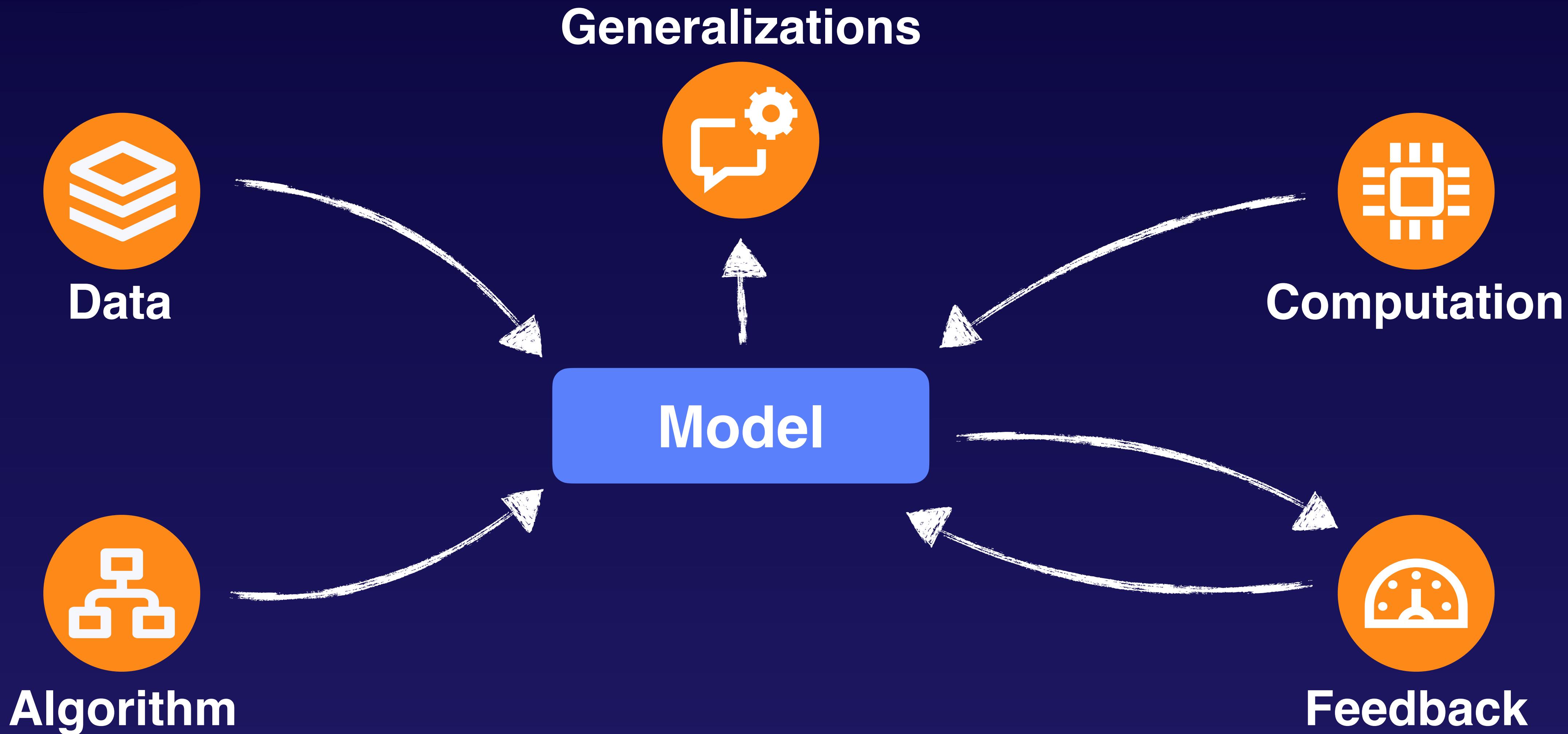
Generalizations



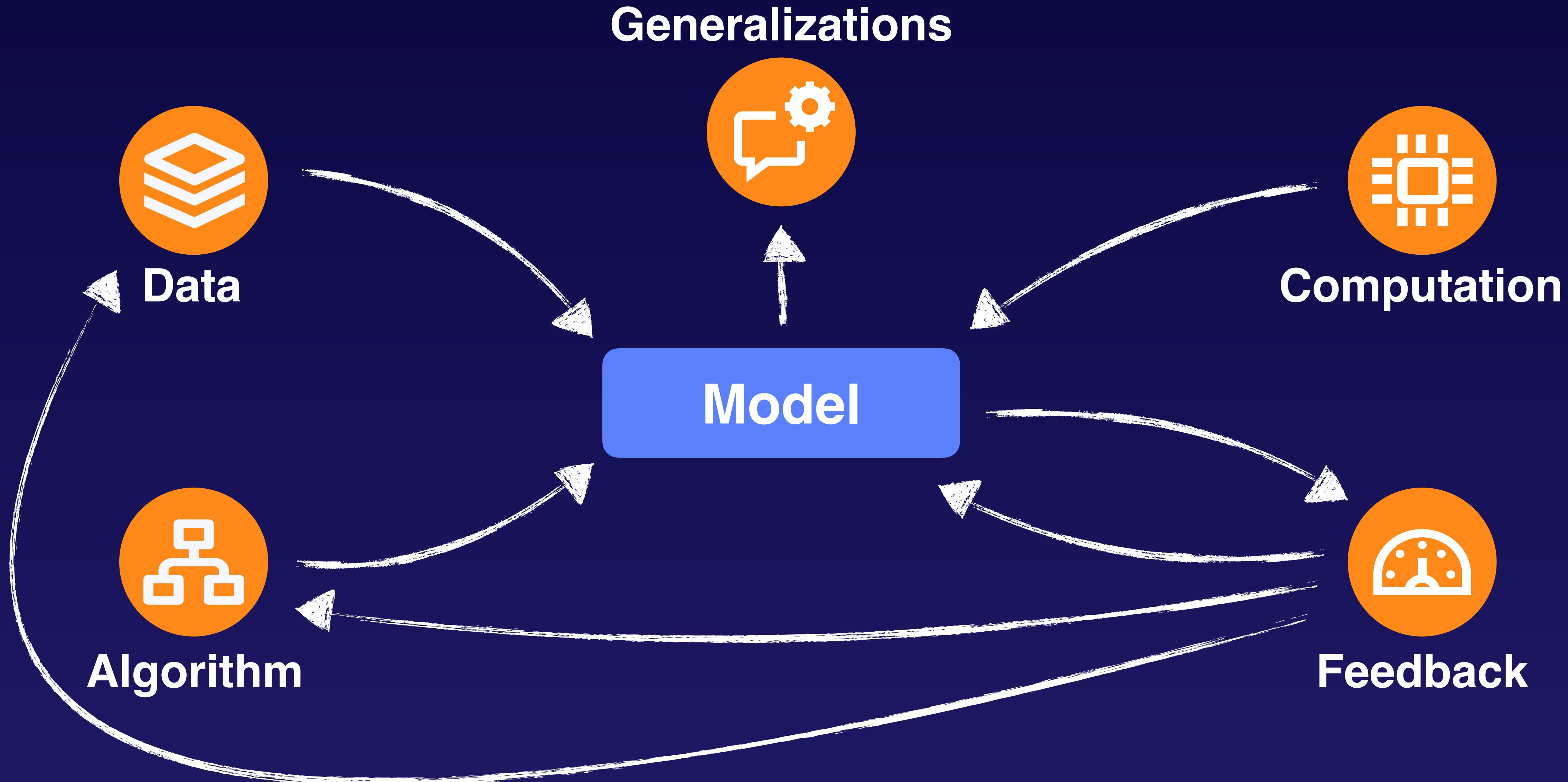
What is a Model?



What is a Model?



What is a Model?



Developing a Good Model

1

What type of generalization are we seeking?

Do I need to forecast a number? Decide whether a customer is most likely to choose Option A or Option B? Detect a quality defect in a machined part?

4

What do we have to work with?

What sort of data accurately and fully captures the inputs and outputs of the target generalization? Do I have enough data? Do I have too much?

2

Do we really need machine learning?

Can simple heuristics handle the job just as well? Can I just program some IF...THEN logic? Will a linear regression formula or a look-up function fulfill the needs?

5

How can I tell if the generalization is working?

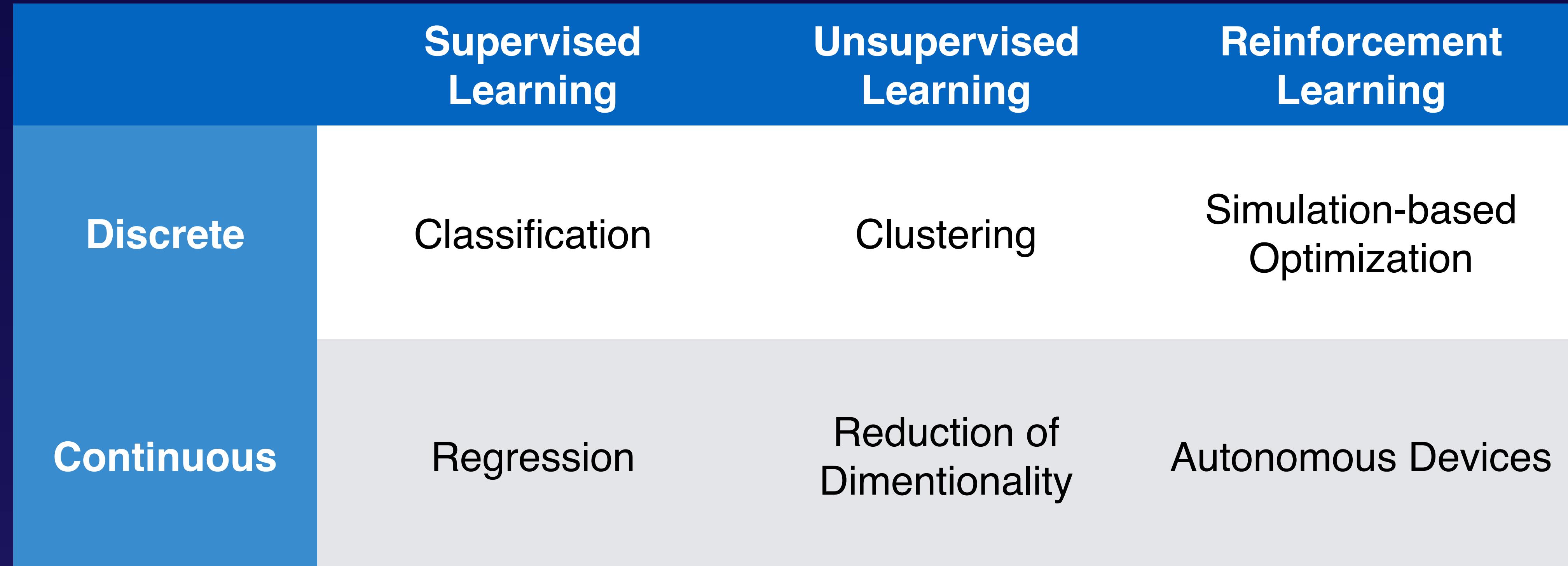
What method can I use to test accuracy and effectiveness? Should my model have a higher sensitivity to false positives or false negatives? How about Accuracy, Recall and Precision?

3

How will my ML generalizations be consumed?

Do I need to return real-time results or can I process the inferences in batch? Will consumers be applications via API call or other systems which will perform additional processing on the data?

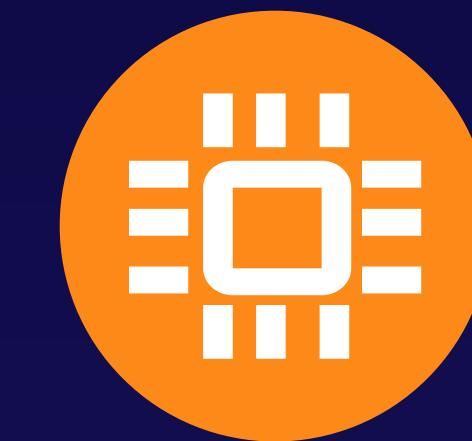
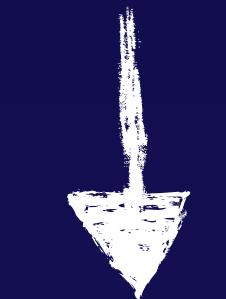
Developing a Good Model



Generalizations

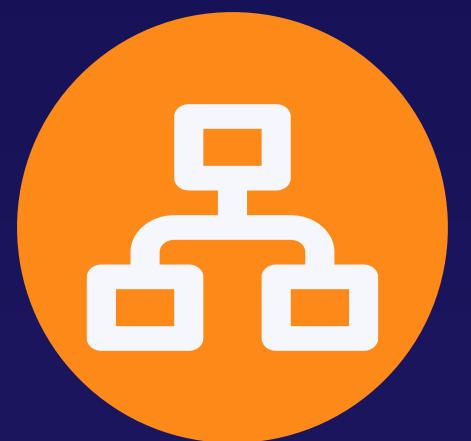


Data



Computation

Model

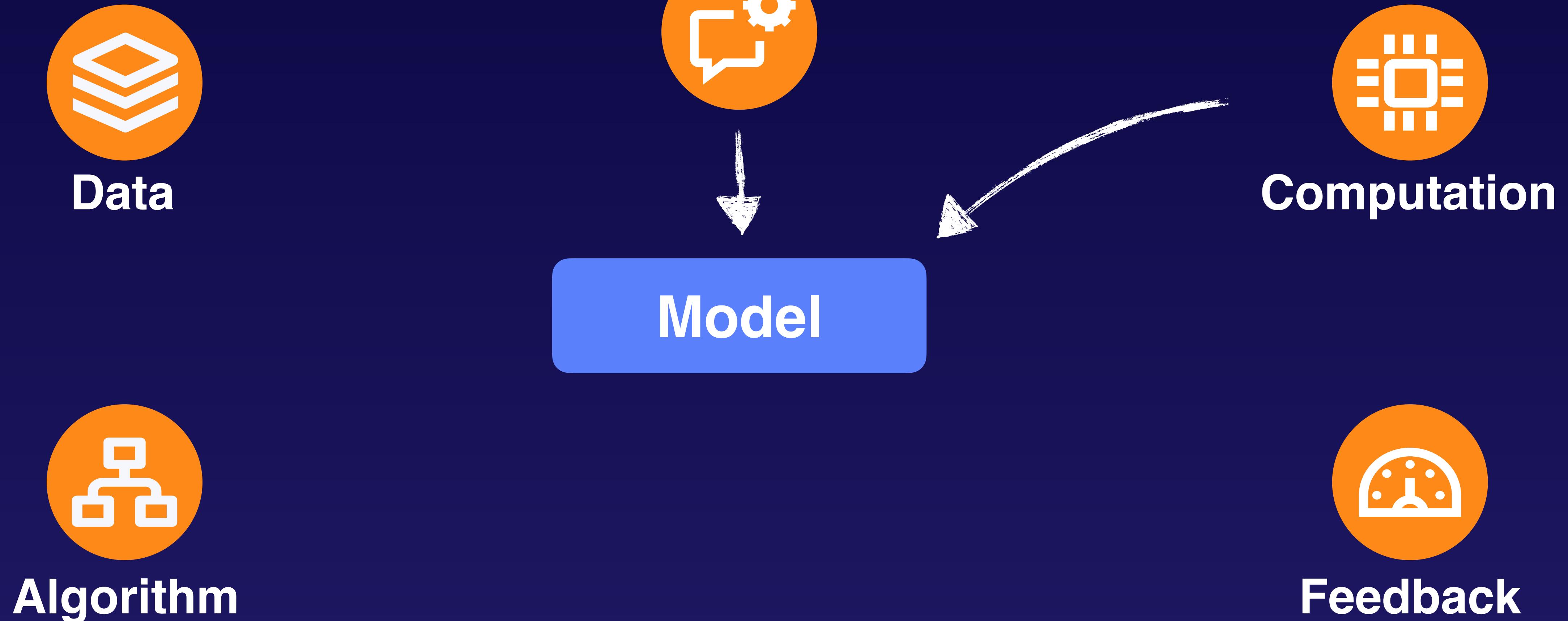


Algorithm

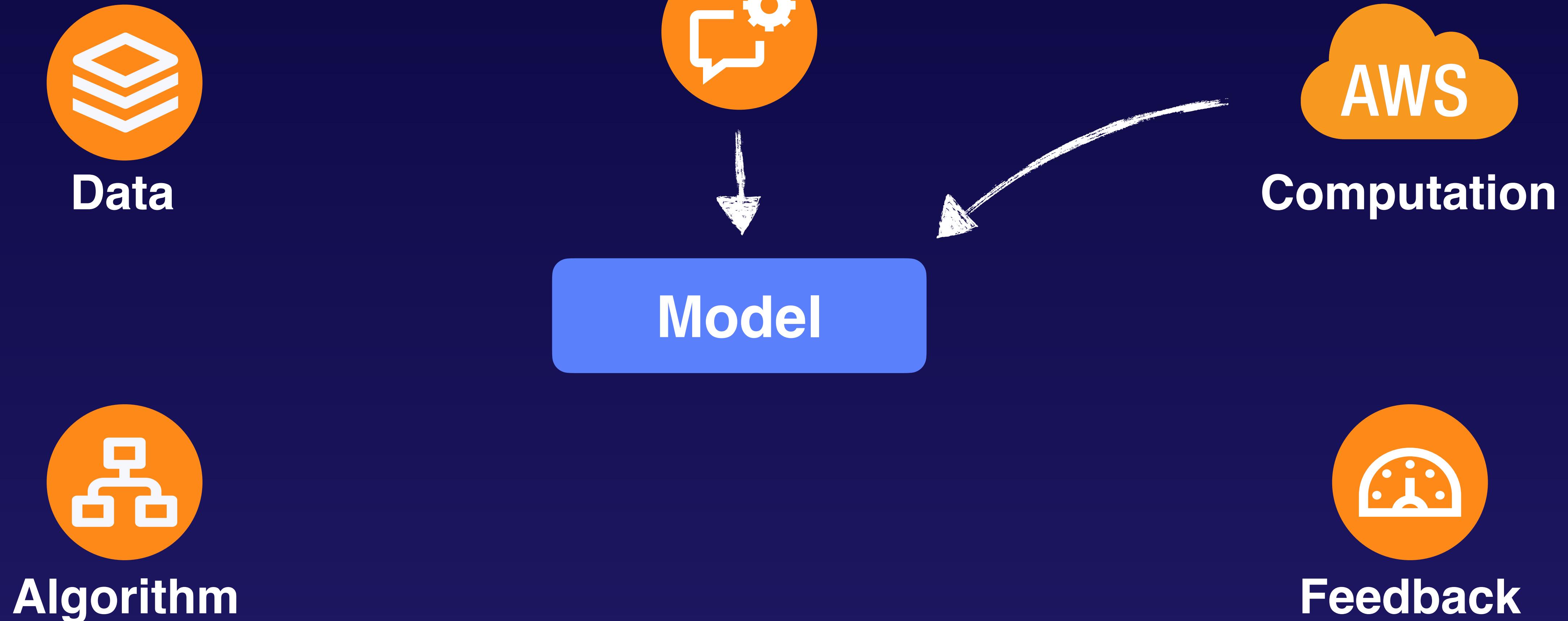


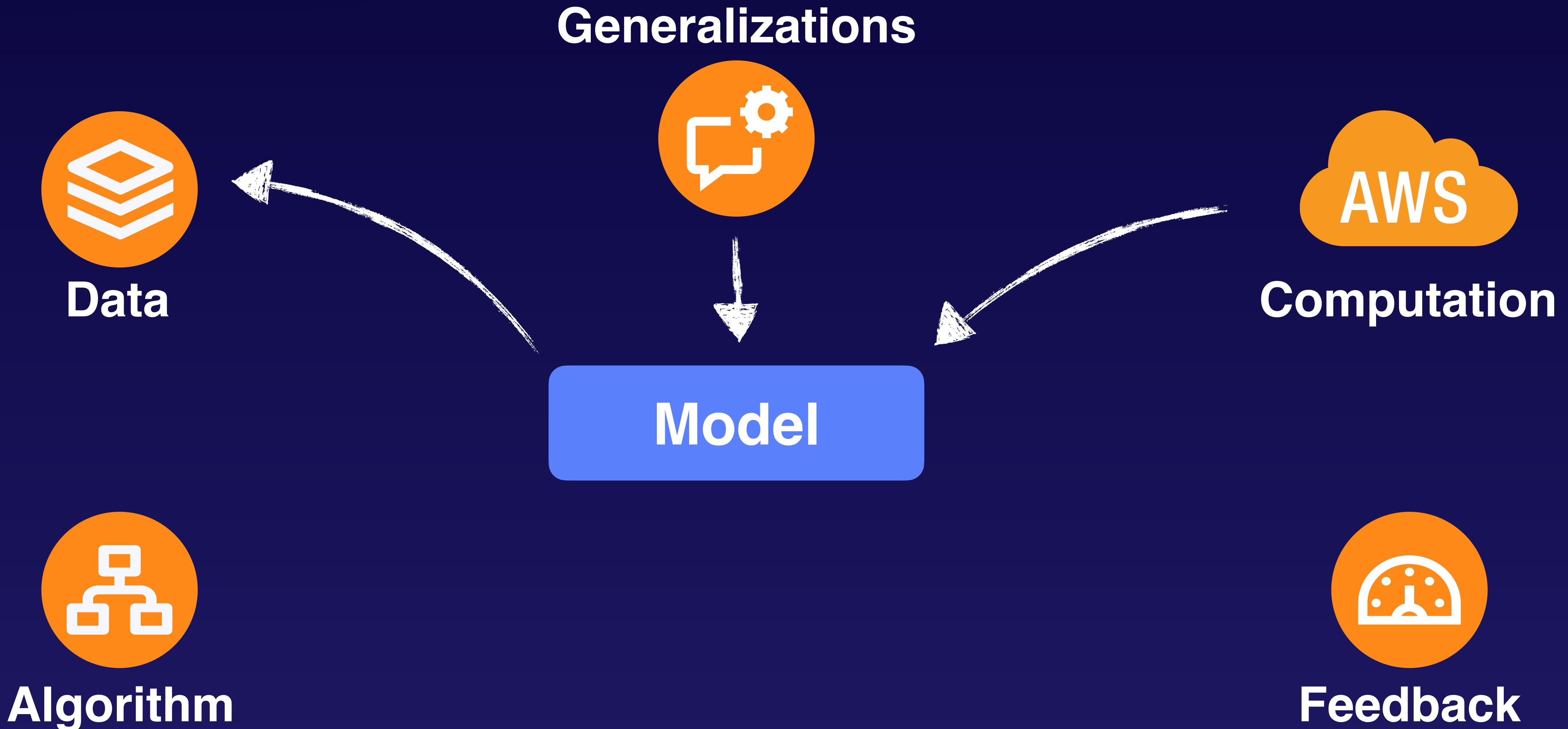
Feedback

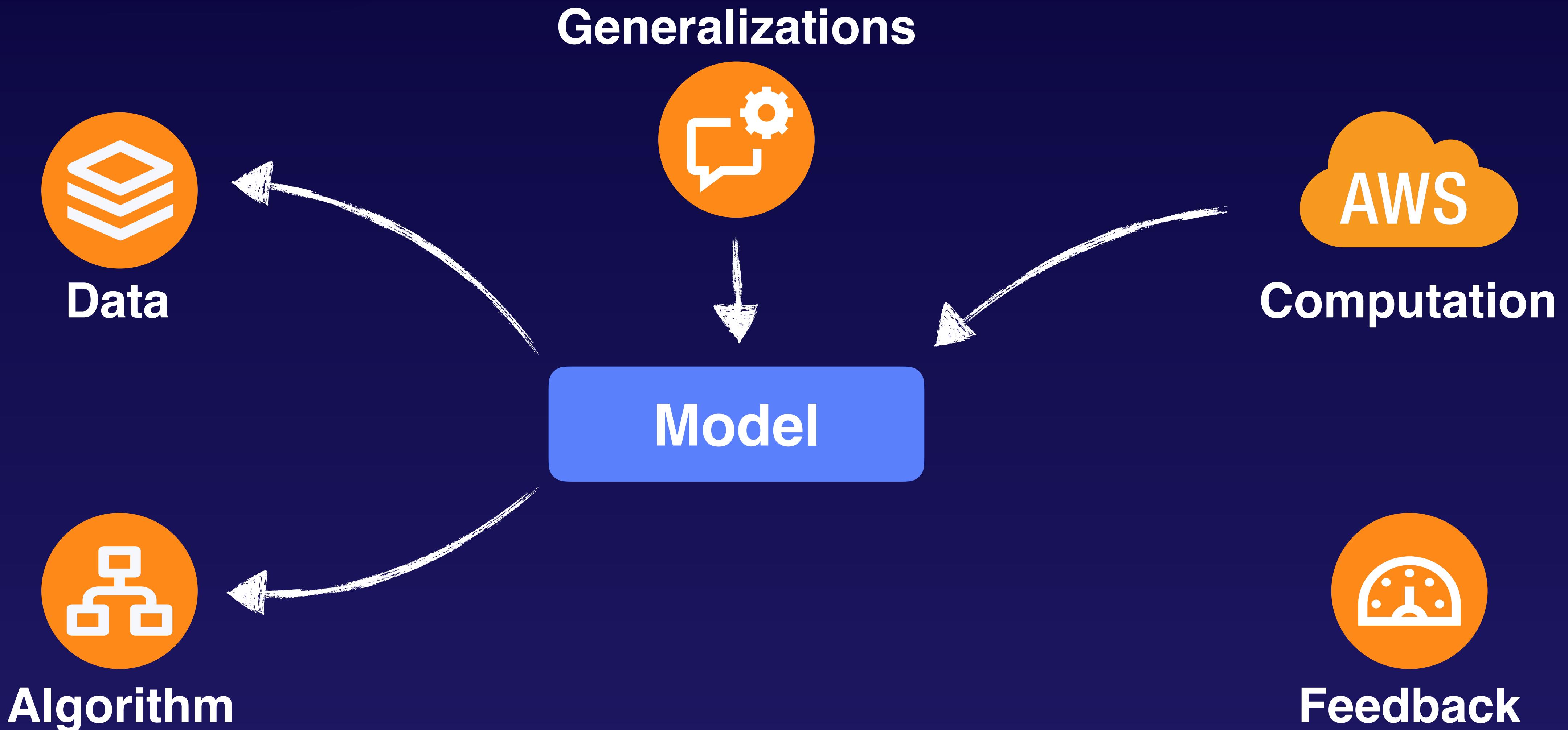
Generalizations

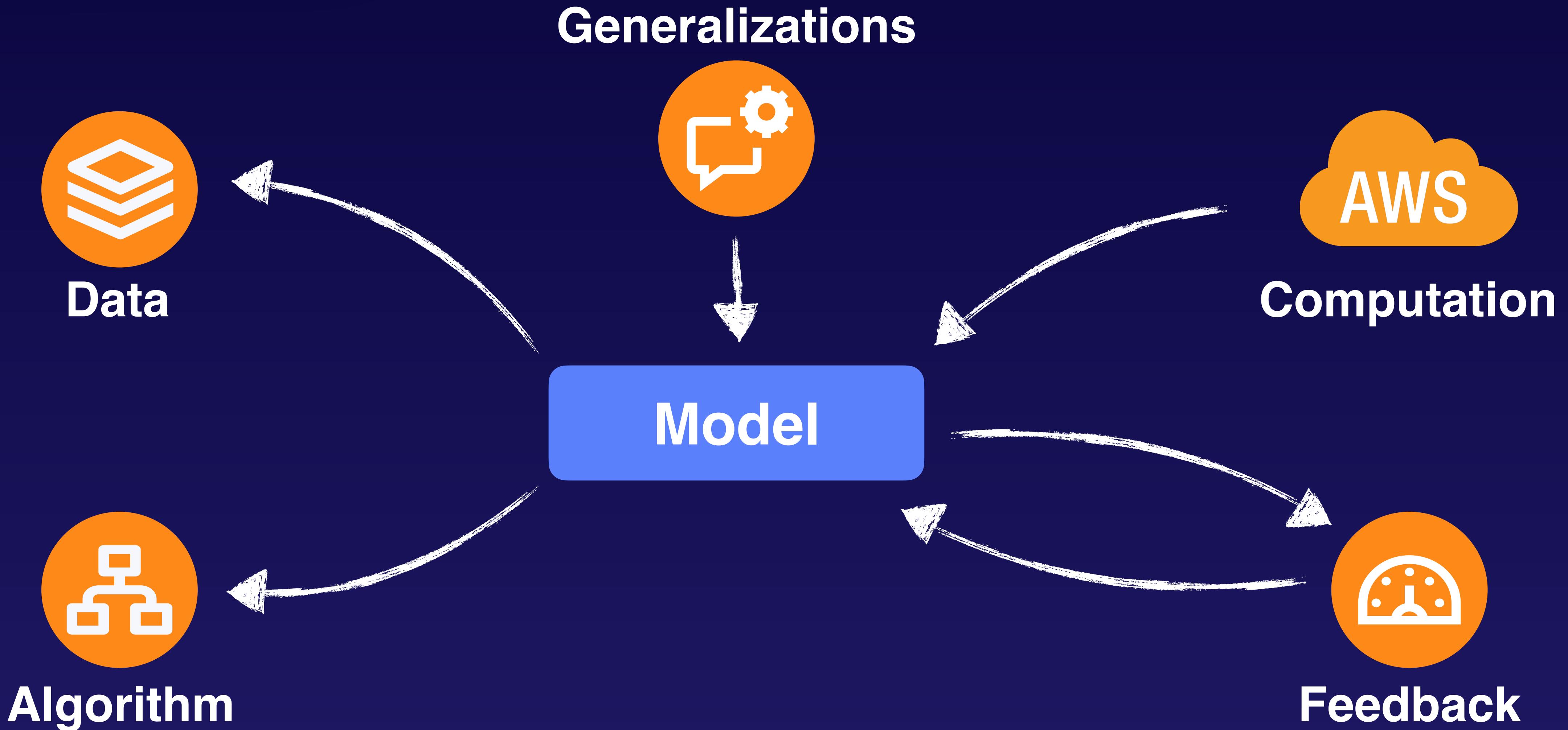


Generalizations



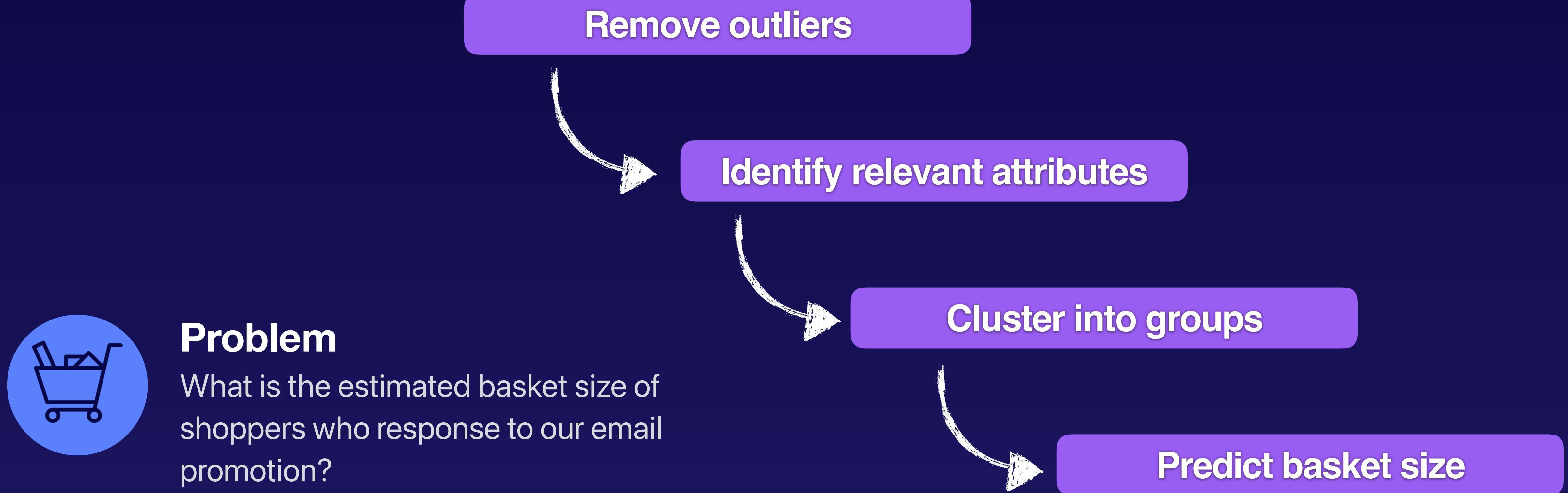






Choosing the Right Approach

| Problem | Approach | Why |
|--|---|--|
| Detect whether a financial transaction is fraud. | Binary Classification | Only two possible outcomes: Fraud or Not Fraud |
| Predict the rate of deceleration of a car when brakes are applied. | Heuristic Approach (No ML Needed!) | Well-known formulas involving speed, inertia and friction to predict this. |
| Determine the most efficient path of surface travel for a robotic lunar rover. | Simulation-based Reinforcement Learning | Must figure out the optimal path via trial, error and improvement. |
| Determine the breed of dog in a photograph. | Multi-Class Classification | Which dog breed is most associated with the picture among many breeds? |



Cascading Algorithms



Problem

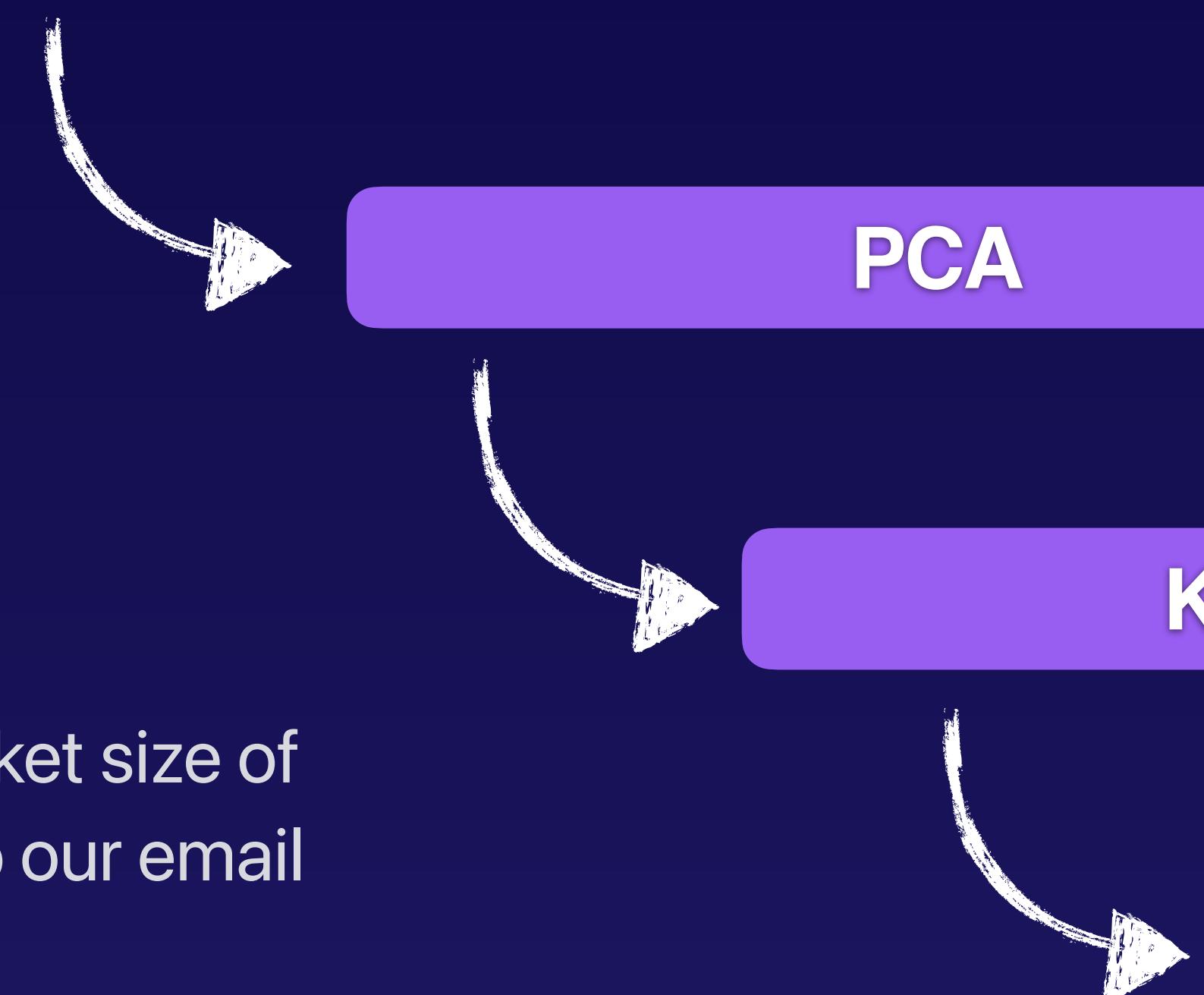
What is the estimated basket size of shoppers who response to our email promotion?

Random Cut Forest

PCA

K-Means

Linear Learner



Confusion Matrix

| | | Actual Outcome | |
|-------------------|-------|----------------------------------|----------------------------------|
| | | TRUE | FALSE |
| Predicted Outcome | TRUE | I predicted correctly! | I was wrong. (False Positive) |
| | FALSE | I was wrong. (False Negative) | I predicted correctly! |

Confusion Matrix



Problem

Is a given financial transaction fraudulent?

| | | Actual Outcome | |
|-------------------|-----------|--------------------------------|--------------------------------|
| | | Fraud | Not Fraud |
| Predicted Outcome | Fraud | Happy Bank. Happy Customer. | |
| | Not Fraud | | Happy Bank. Happy Customer. |

Confusion Matrix



Problem

Is a given financial transaction fraudulent?

| | | Actual Outcome | |
|-------------------|-----------|--------------------------------|--------------------------------|
| | | Fraud | Not Fraud |
| Predicted Outcome | Fraud | Happy Bank. Happy Customer. | |
| | Not Fraud | Angry Bank. Angry Customer. | Happy Bank. Happy Customer. |

Confusion Matrix



Problem

Is a given financial transaction fraudulent?

| | | Actual Outcome | |
|-------------------|-----------|--------------------------------|--------------------------------|
| | | Fraud | Not Fraud |
| Predicted Outcome | Fraud | Happy Bank. Happy Customer. | Happy Bank. Angry Customer. |
| | Not Fraud | Angry Bank. Angry Customer. | Happy Bank. Happy Customer. |

Confusion Matrix



Problem

Is a given financial transaction fraudulent?

| | | Actual Outcome | |
|-------------------|-----------|--------------------------------|--------------------------------|
| | | Fraud | Not Fraud |
| Predicted Outcome | Fraud | Happy Bank. Happy Customer. | Happy Bank. Angry Customer. |
| | Not Fraud | Angry Bank. Angry Customer. | Happy Bank. Happy Customer. |

Confusion Matrix



Problem

Is a given financial transaction fraudulent?

| | | Actual Outcome | |
|-------------------|-----------|--------------------|---------------|
| | | Fraud | Not Fraud |
| Predicted Outcome | Fraud | No Money Loss | No Money Loss |
| | Not Fraud | Money Lost! | No Money Loss |

Confusion Matrix



Problem

Is a given financial transaction fraudulent?

| | | Actual Outcome | |
|-------------------|-----------|--------------------------------|--------------------------------|
| | | Fraud | Not Fraud |
| Predicted Outcome | Fraud | Happy Bank. Happy Customer. | Happy Bank. Angry Customer. |
| | Not Fraud | Angry Bank. Angry Customer. | Happy Bank. Happy Customer. |

Confusion Matrix



Problem

Is a given financial transaction fraudulent?

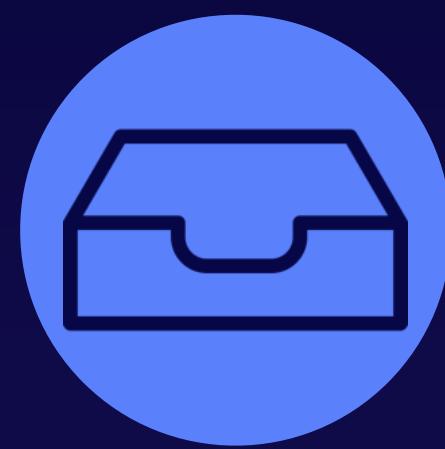
Evaluation Approach

The Bank is likely ok with more false positives than false negatives as that further reduces their exposure to fraud.

We'd closely look at Recall.

| | | Actual Outcome | |
|-------------------|-----------|--------------------------------|--------------------------------|
| | | Fraud | Not Fraud |
| Predicted Outcome | Fraud | Happy Bank. Happy Customer. | Happy Bank. Angry Customer. |
| | Not Fraud | Angry Bank. Angry Customer. | Happy Bank. Happy Customer. |

Confusion Matrix

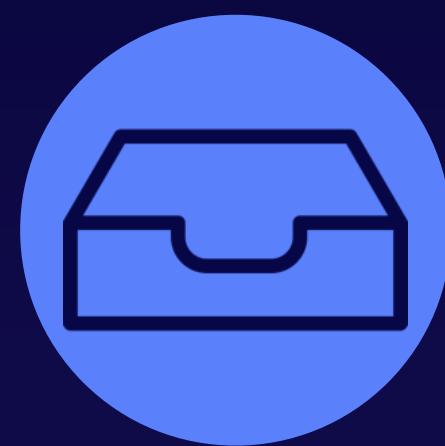


Problem

Is this email spam?

| | | Actual Outcome | |
|-------------------|----------|--------------------|--------------------------------|
| | | Spam | Not Spam |
| Predicted Outcome | Spam | Spam is blocked. | Legitimate emails are blocked. |
| | Not Spam | Spam gets through. | Legitimate emails get through. |

Confusion Matrix

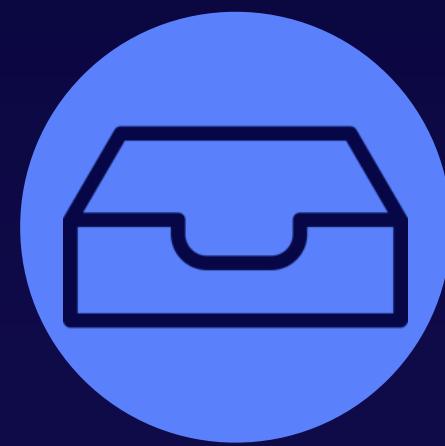


Problem

Is this email spam?

| | | Actual Outcome | |
|-------------------|----------|--------------------|--------------------------------|
| | | Spam | Not Spam |
| Predicted Outcome | Spam | Spam is blocked. | Legitimate emails are blocked. |
| | Not Spam | Spam gets through. | Legitimate emails get through. |

Confusion Matrix



Problem

Is this email spam?

Evaluation Approach

Set the evaluation approach to be more error on the side of caution to let spam through.

We'd watch the Precision of the model closely.

| | | Actual Outcome | |
|-------------------|----------|--------------------|--------------------------------|
| | | Spam | Not Spam |
| Predicted Outcome | Spam | Spam is blocked. | Legitimate emails are blocked. |
| | Not Spam | Spam gets through. | Legitimate emails get through. |