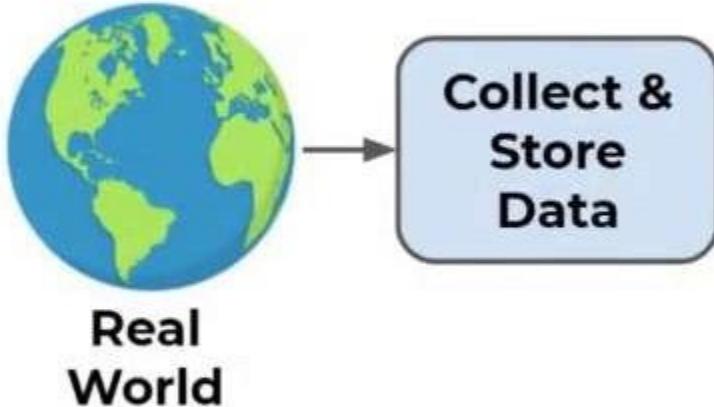


Supervised Machine Learning Process

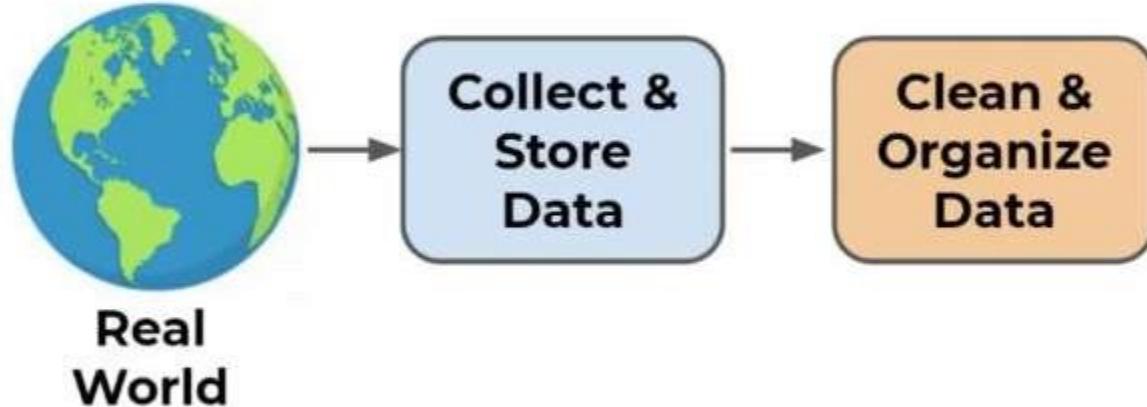
Machine Learning

- Machine Learning Pathway



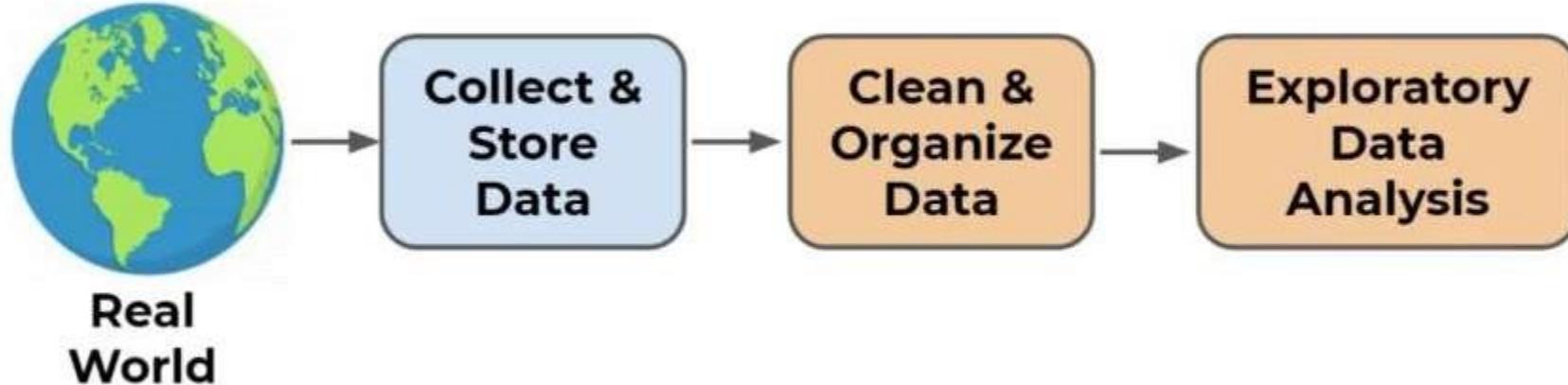
Machine Learning

- Machine Learning Pathway



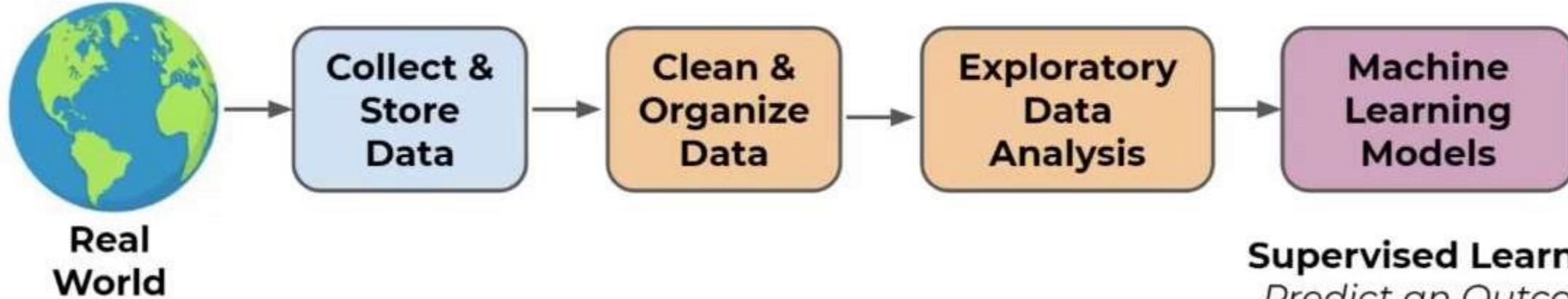
Machine Learning

- Machine Learning Pathway



Machine Learning

- Machine Learning Pathway

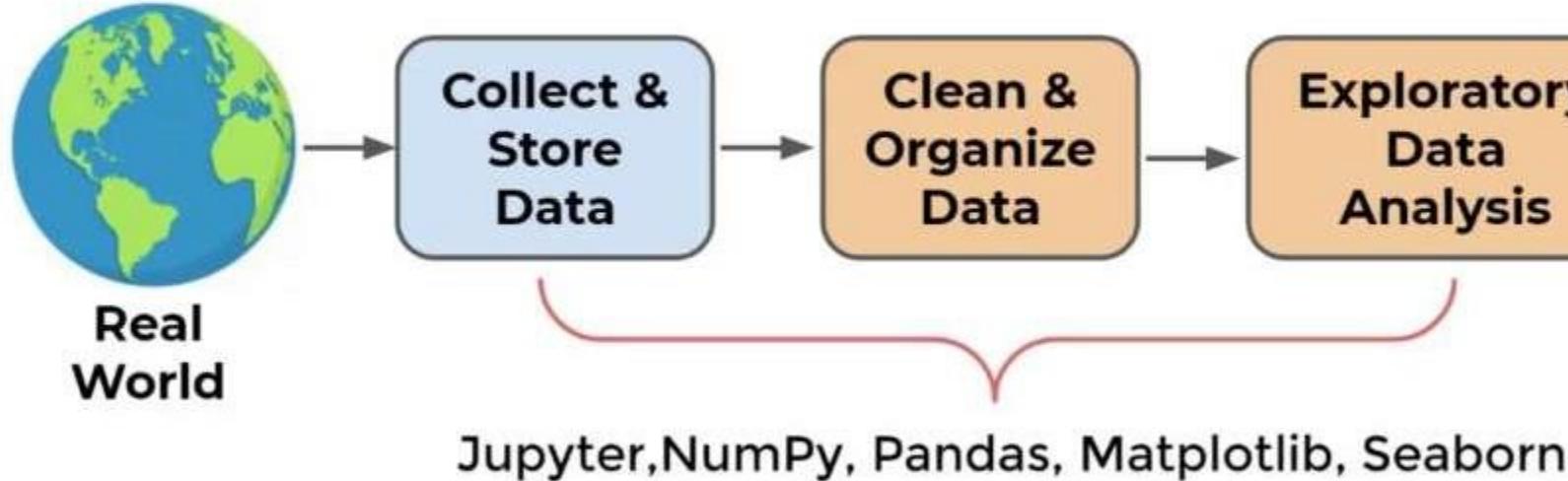


Supervised Learning:
Predict an Outcome

Unsupervised Learning:
Discover Patterns in Data

Machine Learning

- Machine Learning Pathway



Jupyter, NumPy, Pandas, Matplotlib, Seaborn

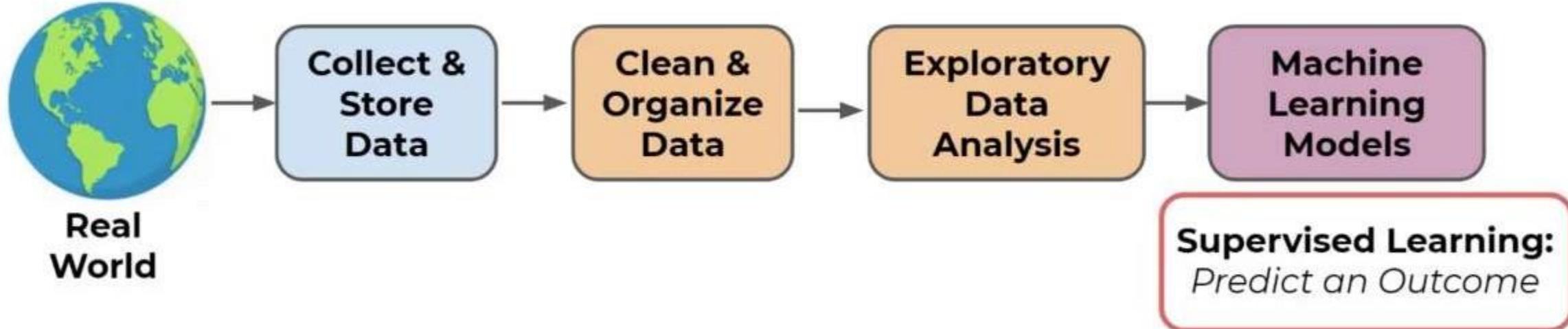
Supervised Learning:
Predict an Outcome

Unsupervised Learning:
Discover Patterns in Data

Scikit-learn

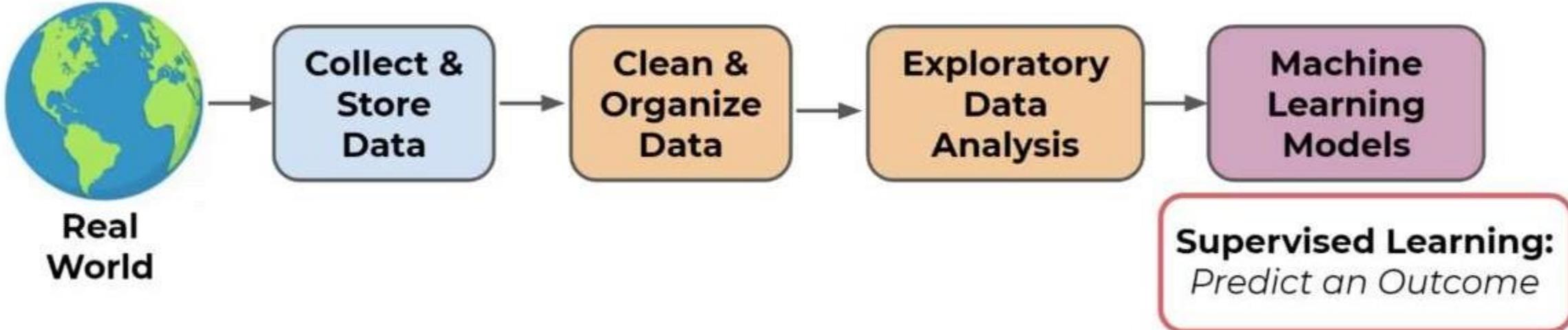
Machine Learning

- Machine Learning Pathway



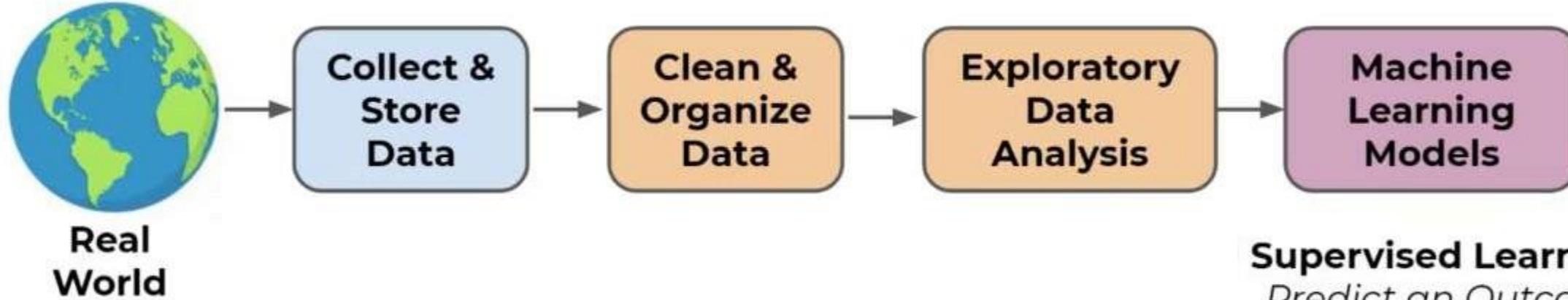
Machine Learning

- ML Process : Supervised Learning Tasks



Machine Learning

- Predict price a house should sell at.



Supervised Learning:
Predict an Outcome

Machine Learning

- **Supervised Machine Learning Process**
- Start with collecting and organizing a data set based on history:

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- **Historical labeled data on previously sold houses.**

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- If a new house comes on the market with a known Area, Bedrooms, and Bathrooms:
Predict what price should it sell at.

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- Data Product:
 - Input house features
 - Output predicted selling price

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

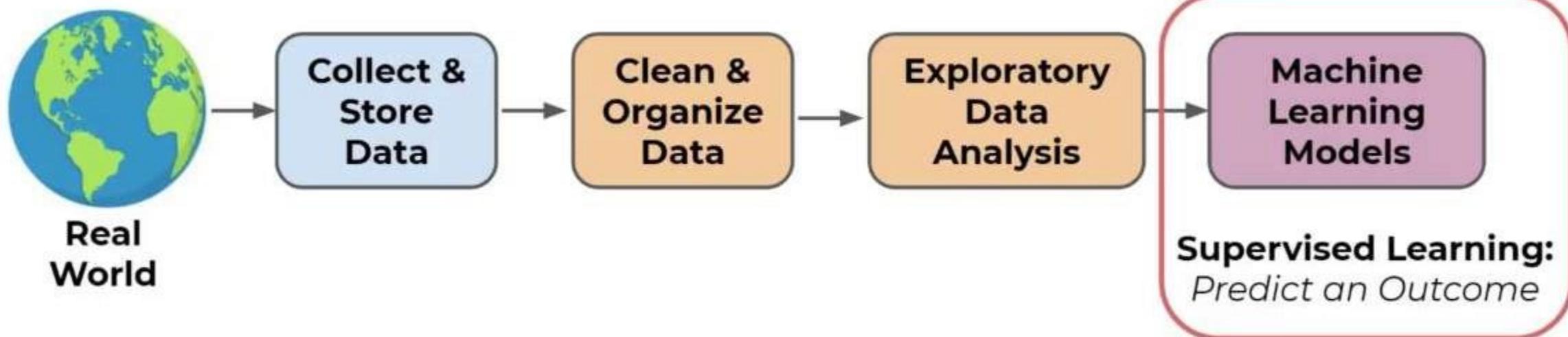
Machine Learning

- Using historical, labeled data predict a future outcome or result.

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- Predict price a house should sell at.



Machine Learning

- Predict price a house should sell at.

Machine
Learning
Models

Supervised Learning:
Predict an Outcome

Machine Learning

- Predict price a house should sell at.

Machine Learning Models

Supervised Learning:
Predict an Outcome

Machine Learning

- Predict price a house should sell at.

Machine Learning Models

Supervised Learning:
Predict an Outcome

Data

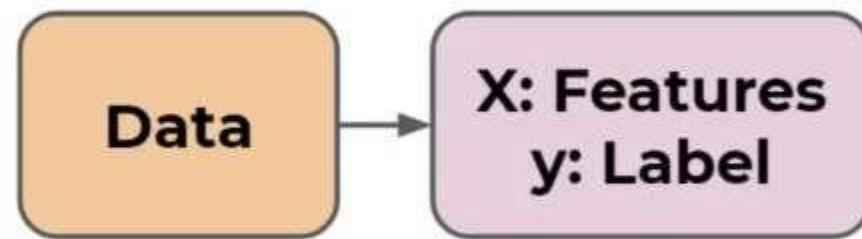
Machine Learning

- Supervised Machine Learning Process

Data

Machine Learning

- Supervised Machine Learning Process



Machine Learning

- **Supervised Machine Learning Process**

Data

X: Features
y: Label

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- **Label** is what we are trying to predict

Data

X: Features
y: Label

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- **Label** is what we are trying to predict

Data

X: Features
y: Label

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- **Features** are known characteristics or components in the data

Data

X: Features
y: Label

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- **Features** are known characteristics or components in the data

Data

X: Features
y: Label

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Machine Learning

- **Features and Label** are identified according to the problem being solved.

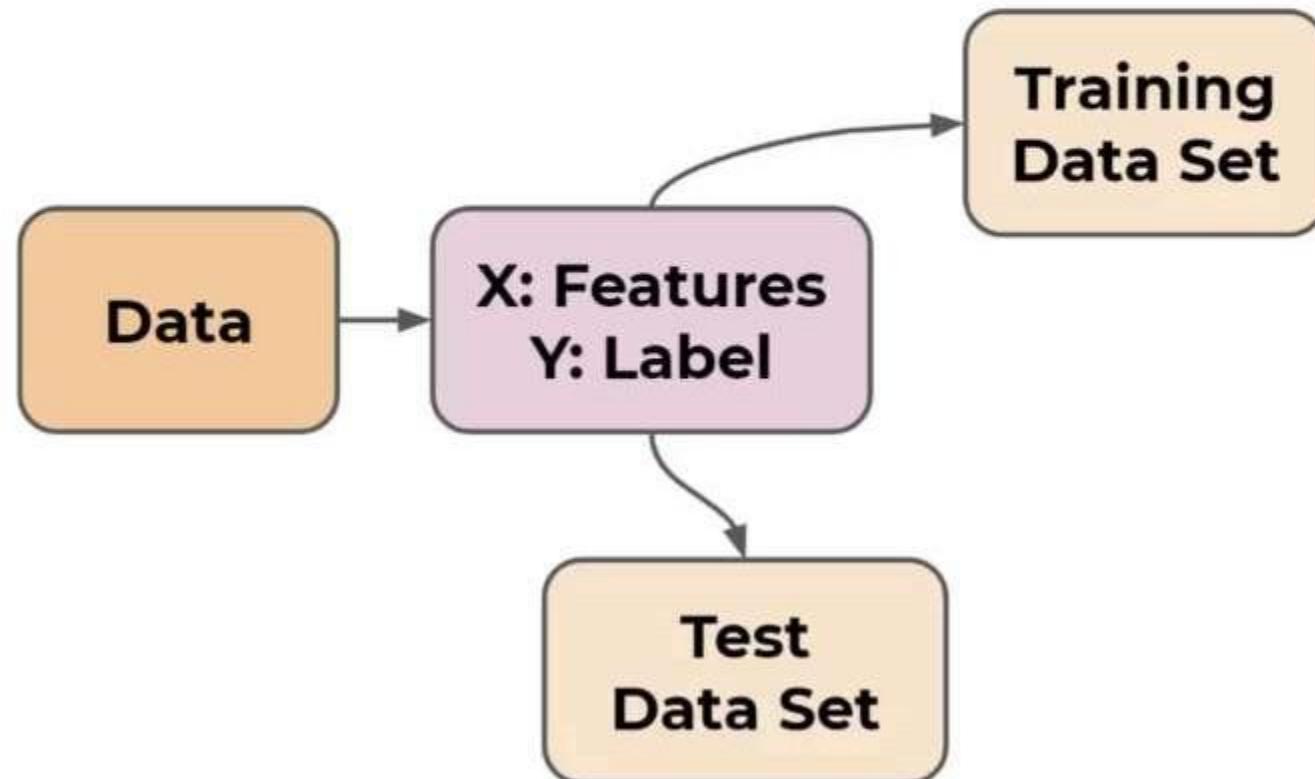
Data

X: Features
y: Label

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

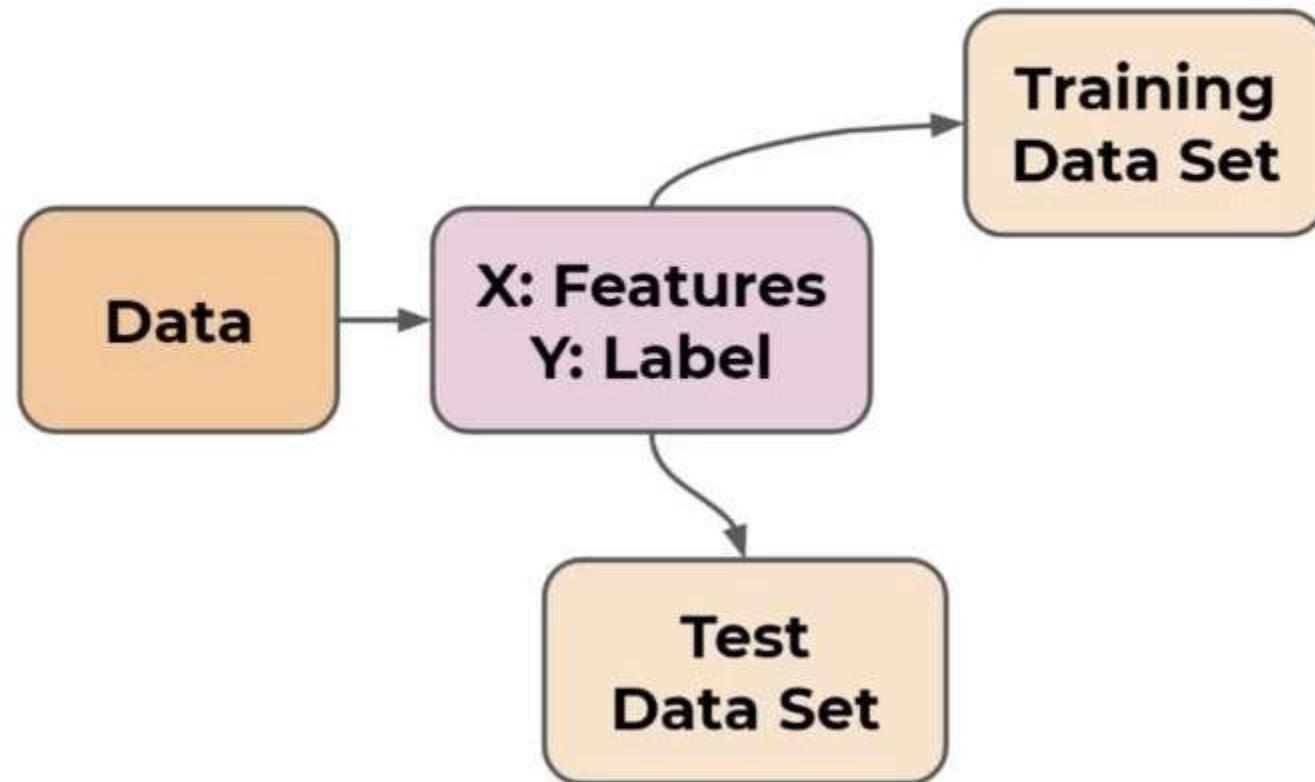
Supervised Machine Learning Process

- Split data into training set and test set



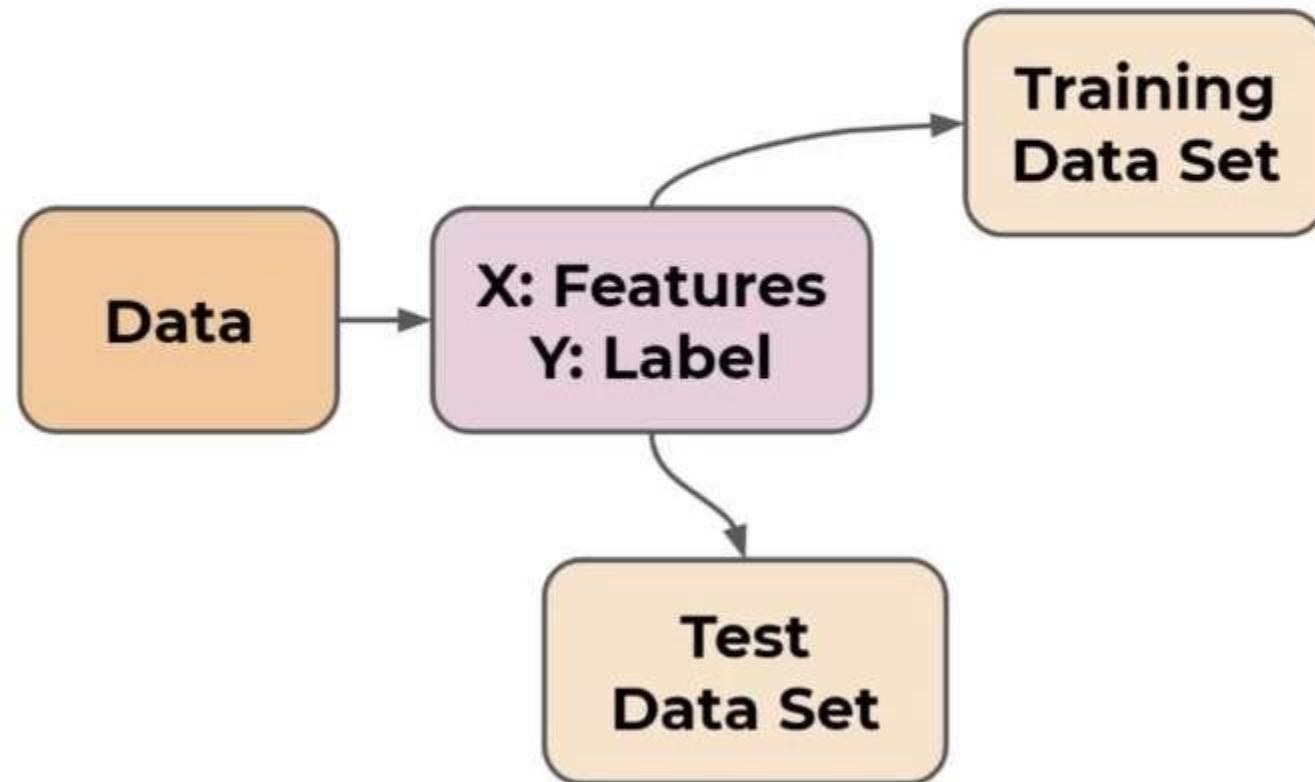
Supervised Machine Learning Process

- Later on we will discuss cross-validation



Supervised Machine Learning Process

- Why perform this split? How to split?



Supervised Machine Learning Process

- Why perform this split? How to split?

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- How would you judge a human realtor's performance?



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000



Supervised Machine Learning Process

- Ask a human realtor to take a look at historical data...



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- Then give her the features of a house and ask her to predict a selling price.



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- But how would you measure how accurate her prediction is? What house should you choose to test on?



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- You can't judge her based on a new house that hasn't sold yet, you don't know its true selling price!



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- You shouldn't judge her on data she's already seen, she could have **memorized** it!



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- Thus the need for a Train/Test split of the data, let's explore further...



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- We already organized the data into **Features (X)** and a **Label (y)**

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- Now we will split this into a training set and a test set:

TRAIN

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- Now we will split this into a training set and a test set:

	Area m ²	Bedrooms	Bathrooms	Price
TRAIN	200	3	2	\$500,000
	190	2	1	\$450,000
	230	3	3	\$650,000
TEST	180	1	1	\$400,000
	210	2	2	\$550,000

Supervised Machine Learning Process

- Notice how we have 4 components

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

X TRAIN **Y TRAIN**

X TEST **Y TEST**



Supervised Machine Learning Process

- Let's go back to fairly testing our human realtor....



Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- Let's go back to fairly testing our human realtor....



TRAIN

TEST

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000
180	1	1	\$400,000
210	2	2	\$550,000

Supervised Machine Learning Process

- Let her study and learn on the training set getting access to both X and y.



TRAIN

Area m ²	Bedrooms	Bathrooms	Price
200	3	2	\$500,000
190	2	1	\$450,000
230	3	3	\$650,000

Supervised Machine Learning Process

- After she has “learned” about the data, we can test her skill on the test set.



TEST

Area m ²	Bedrooms	Bathrooms
180	1	1
210	2	2



Supervised Machine Learning Process

- Provide only the X test data and ask for her predictions for the sell price.



TEST

Area m ²	Bedrooms	Bathrooms
180	1	1
210	2	2

Supervised Machine Learning Process

- This is new data she has never seen before!
She has also never seen the real sold price.



TEST

Area m ²	Bedrooms	Bathrooms
180	1	1
210	2	2

Supervised Machine Learning Process

- Ask for predictions per data point.



Predictions	Area m ²	Bedrooms	Bathrooms
\$410,000	180	1	1
\$540,000	210	2	2

Supervised Machine Learning Process

- Then bring back the original prices.



Predictions	Area m ²	Bedrooms	Bathrooms	Price
\$410,000	180	1	1	\$400,000
\$540,000	210	2	2	\$550,000

Supervised Machine Learning Process

- Finally compare predictions against true test price.



Predictions	Price
\$410,000	\$400,000
\$540,000	\$550,000

Supervised Machine Learning Process

- This is often labeled as \hat{y} compared again y



\hat{y}	y
Predictions	Price
\$410,000	\$400,000
\$540,000	\$550,000

Supervised Machine Learning Process

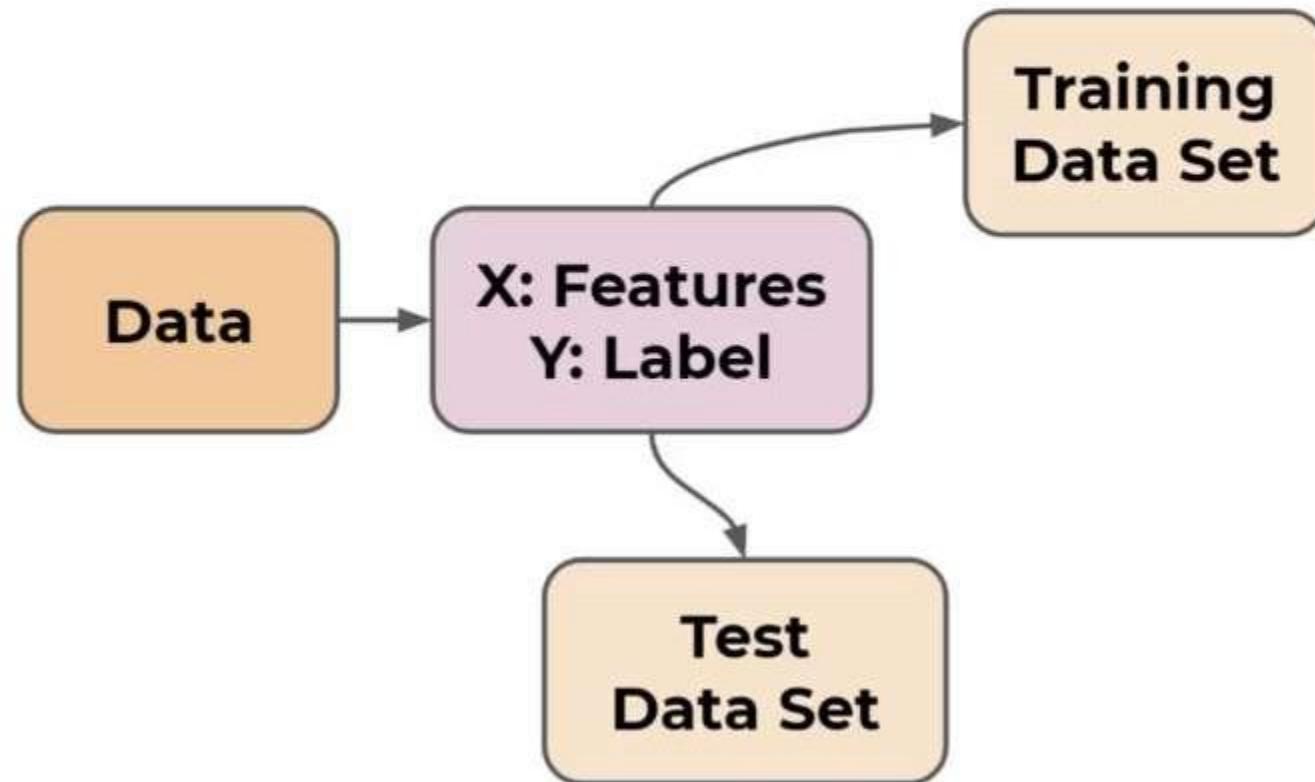
- Later on we will discuss the many methods of evaluating this performance!



Predictions	Price
\$410,000	\$400,000
\$540,000	\$550,000

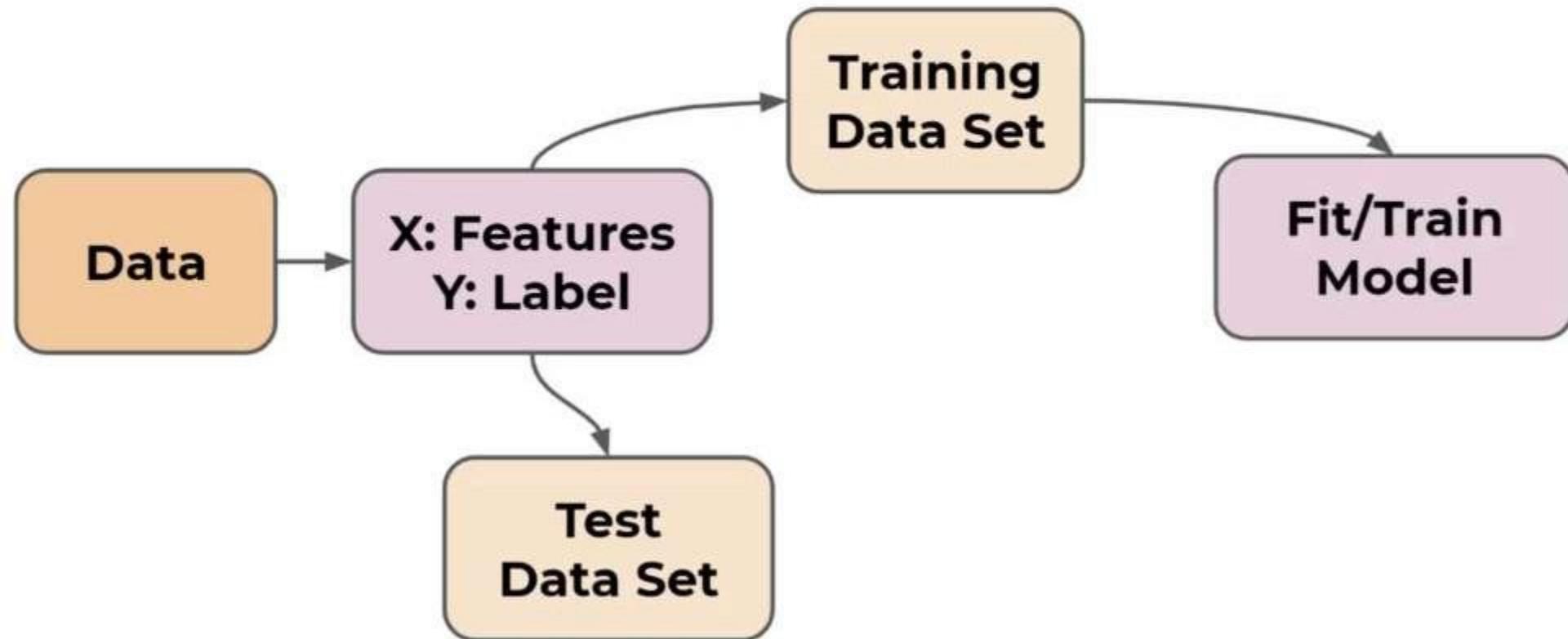
Supervised Machine Learning Process

- Split Data



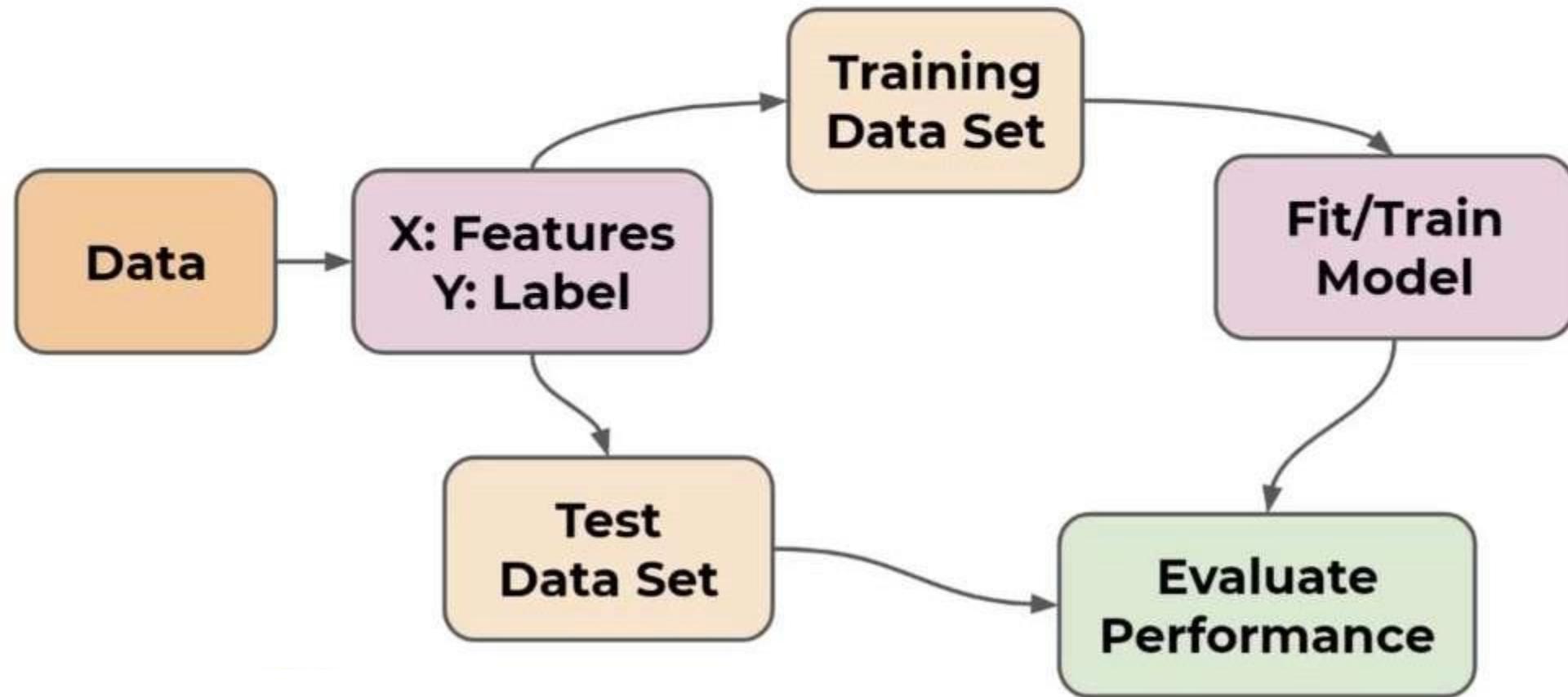
Supervised Machine Learning Process

- Split Data, Fit on Train Data



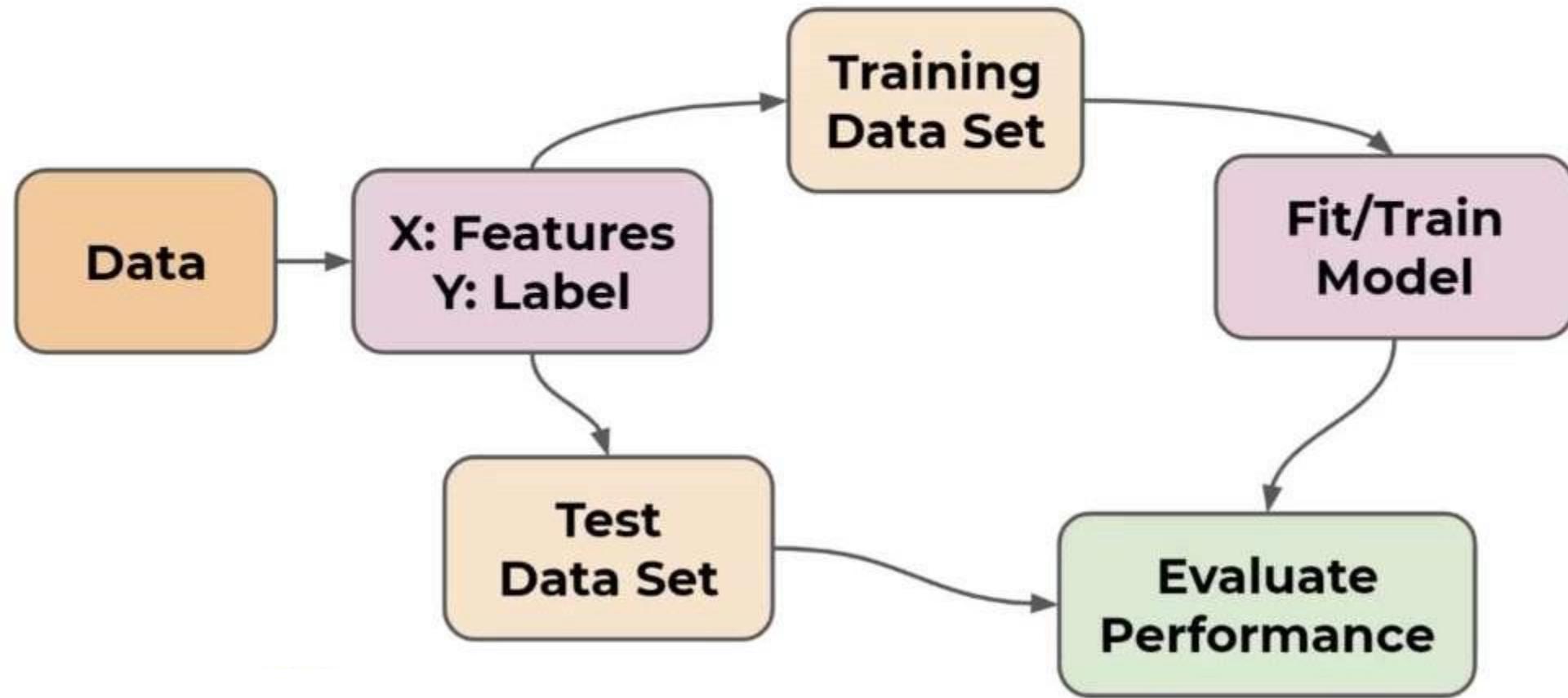
Supervised Machine Learning Process

- Split Data, Fit on Train Data, Evaluate Model



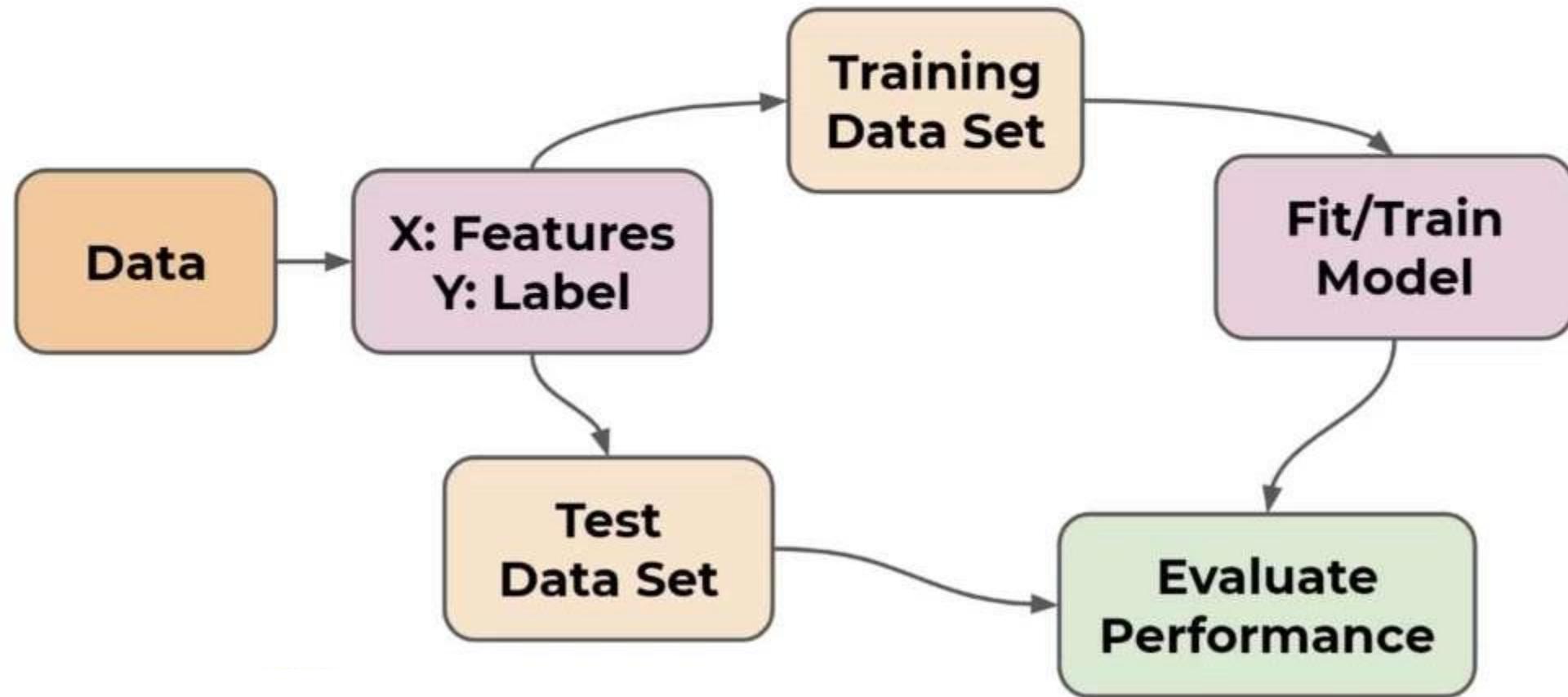
Supervised Machine Learning Process

- What happens if performance isn't great?



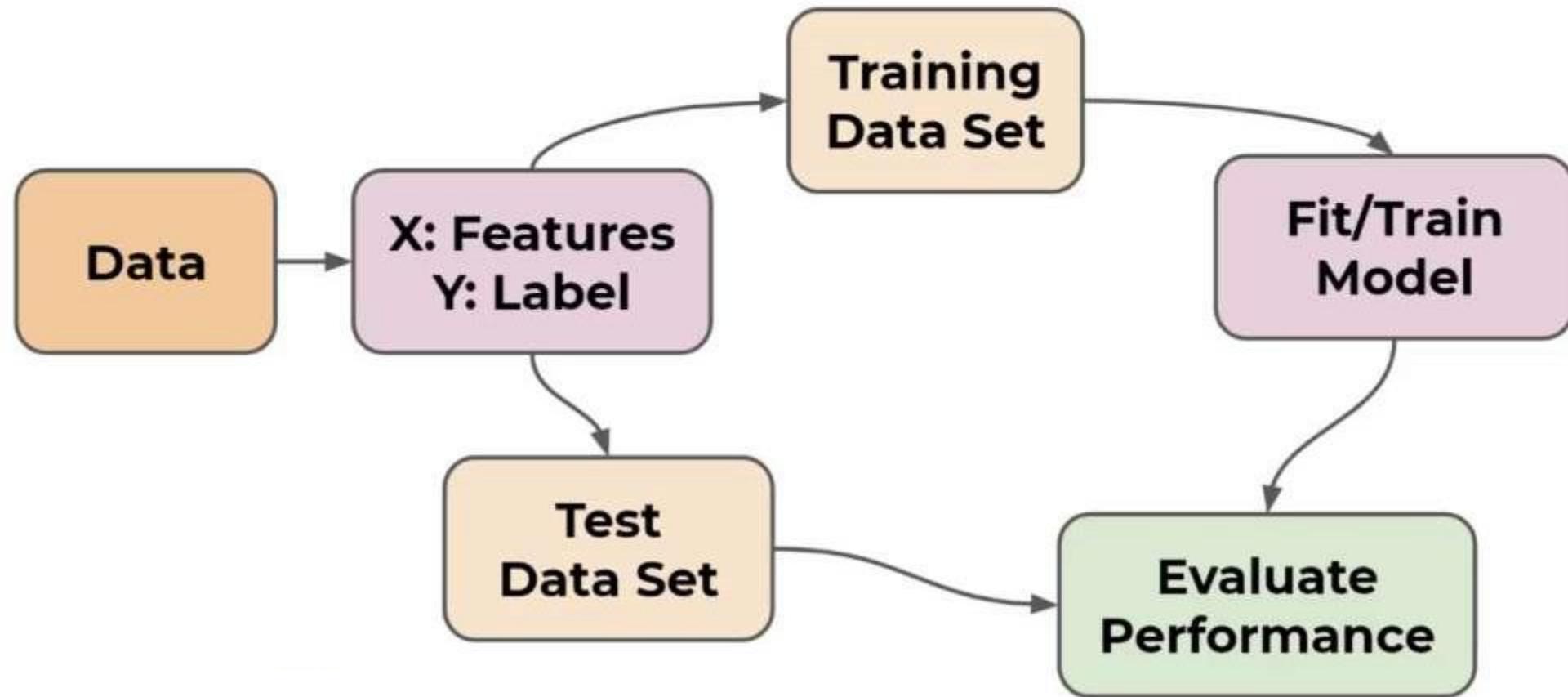
Supervised Machine Learning Process

- We can adjust model **hyperparameters**



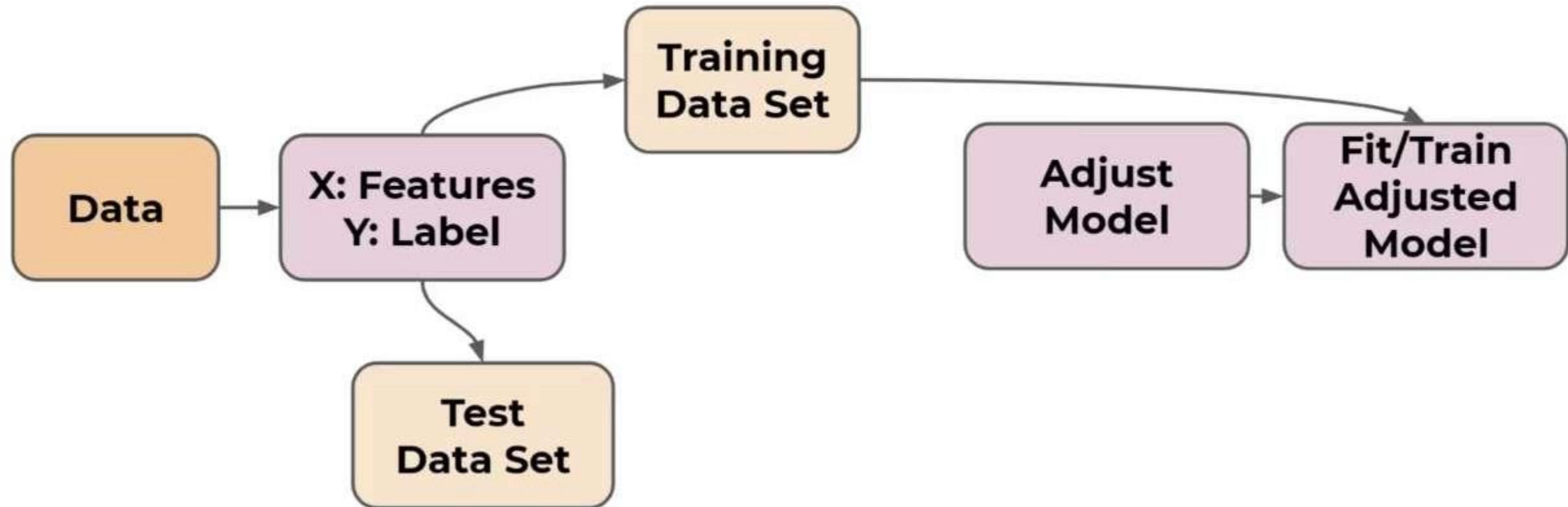
Supervised Machine Learning Process

- Many algorithms have adjustable values



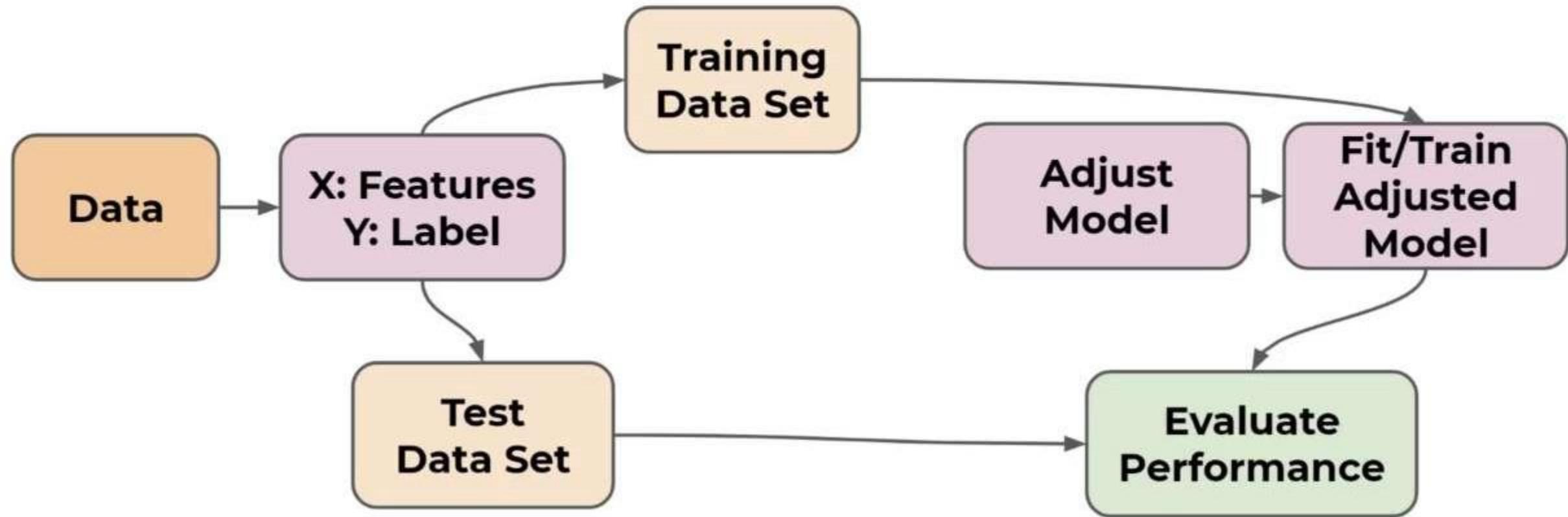
Supervised Machine Learning Process

- Many algorithms have adjustable values



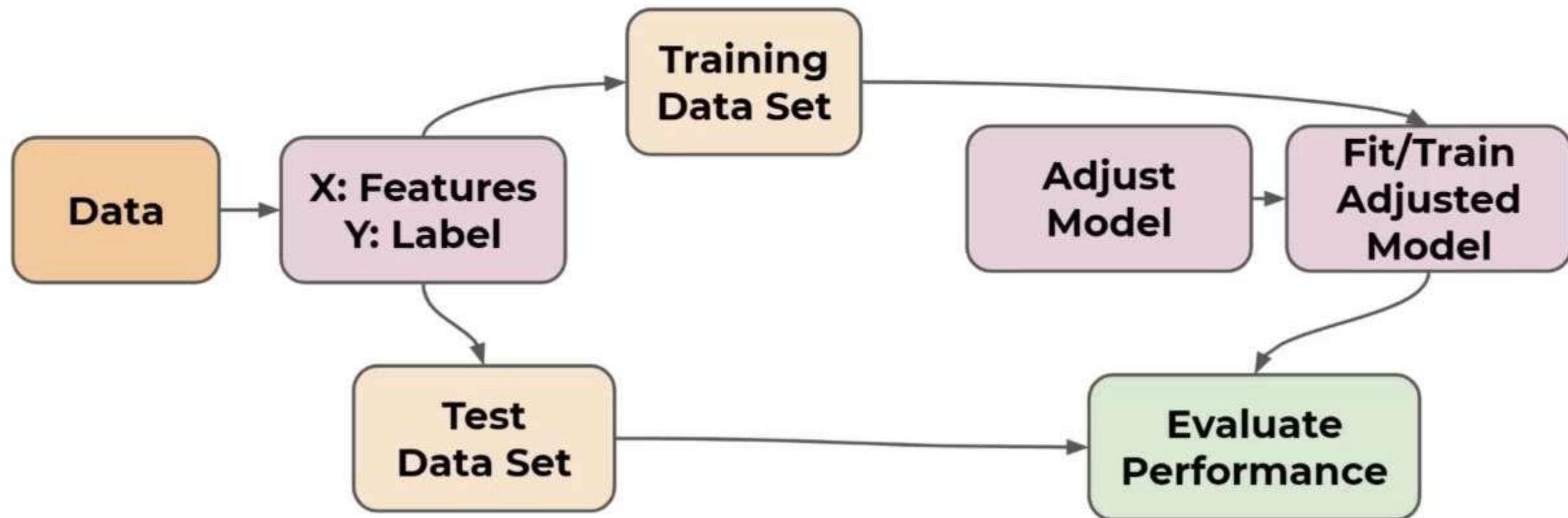
Supervised Machine Learning Process

- Evaluate adjusted model



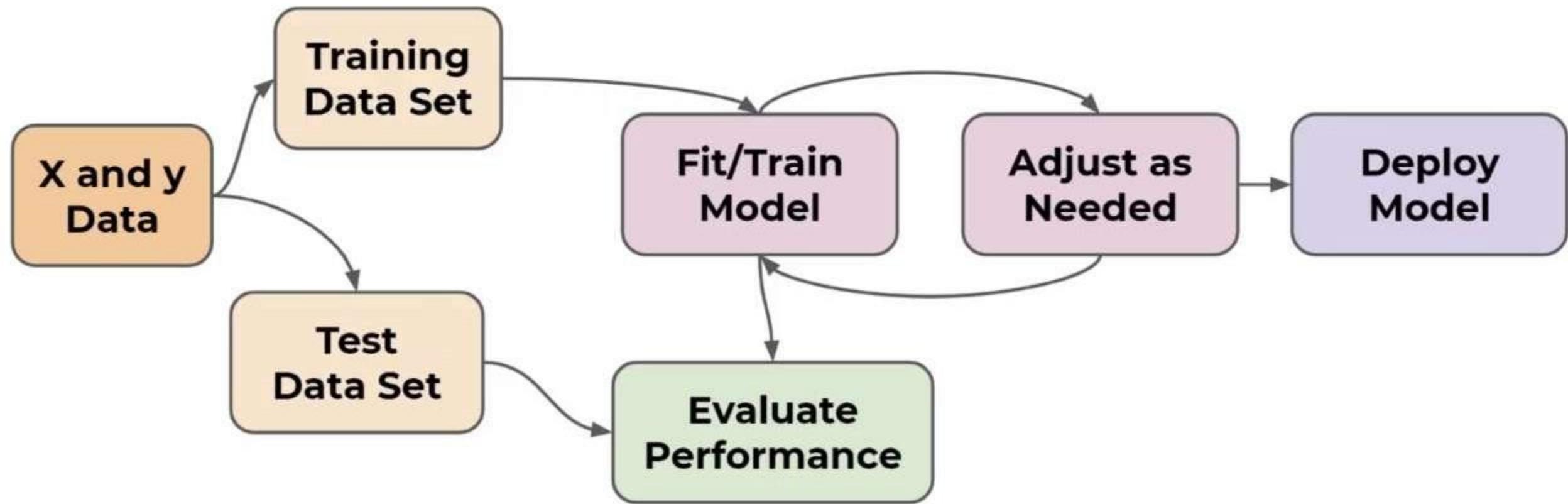
Supervised Machine Learning Process

- Can repeat this process as necessary



Supervised Machine Learning Process

- Full and Simplified Process



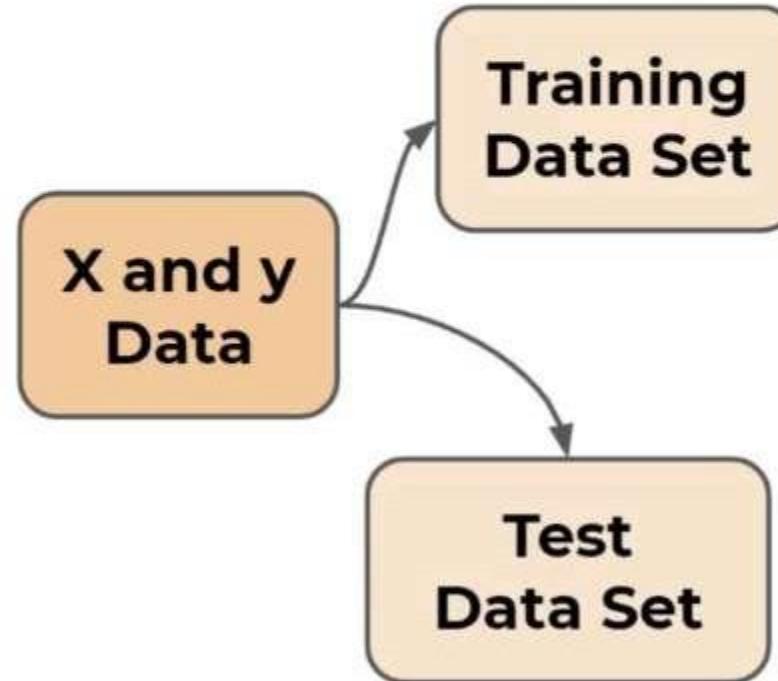
Supervised Machine Learning Process

- Get X and y data

X and y
Data

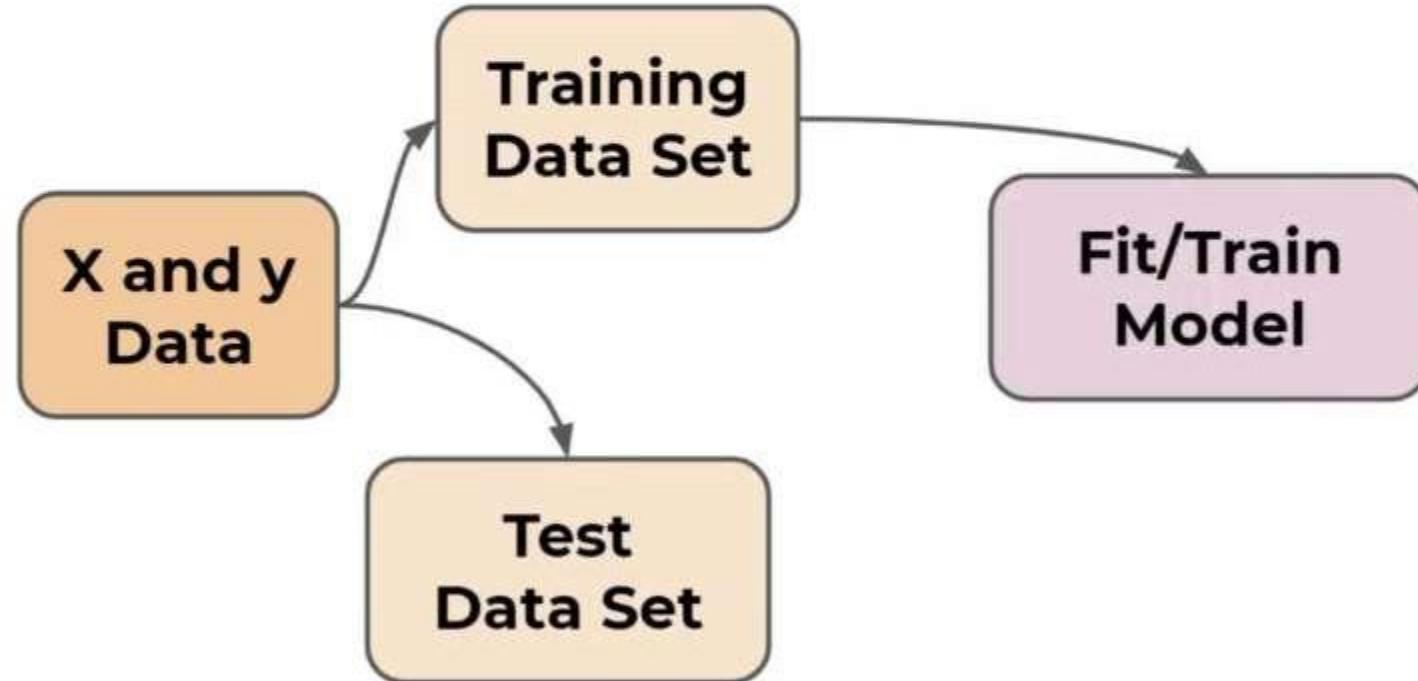
Supervised Machine Learning Process

- Split data for evaluation purposes



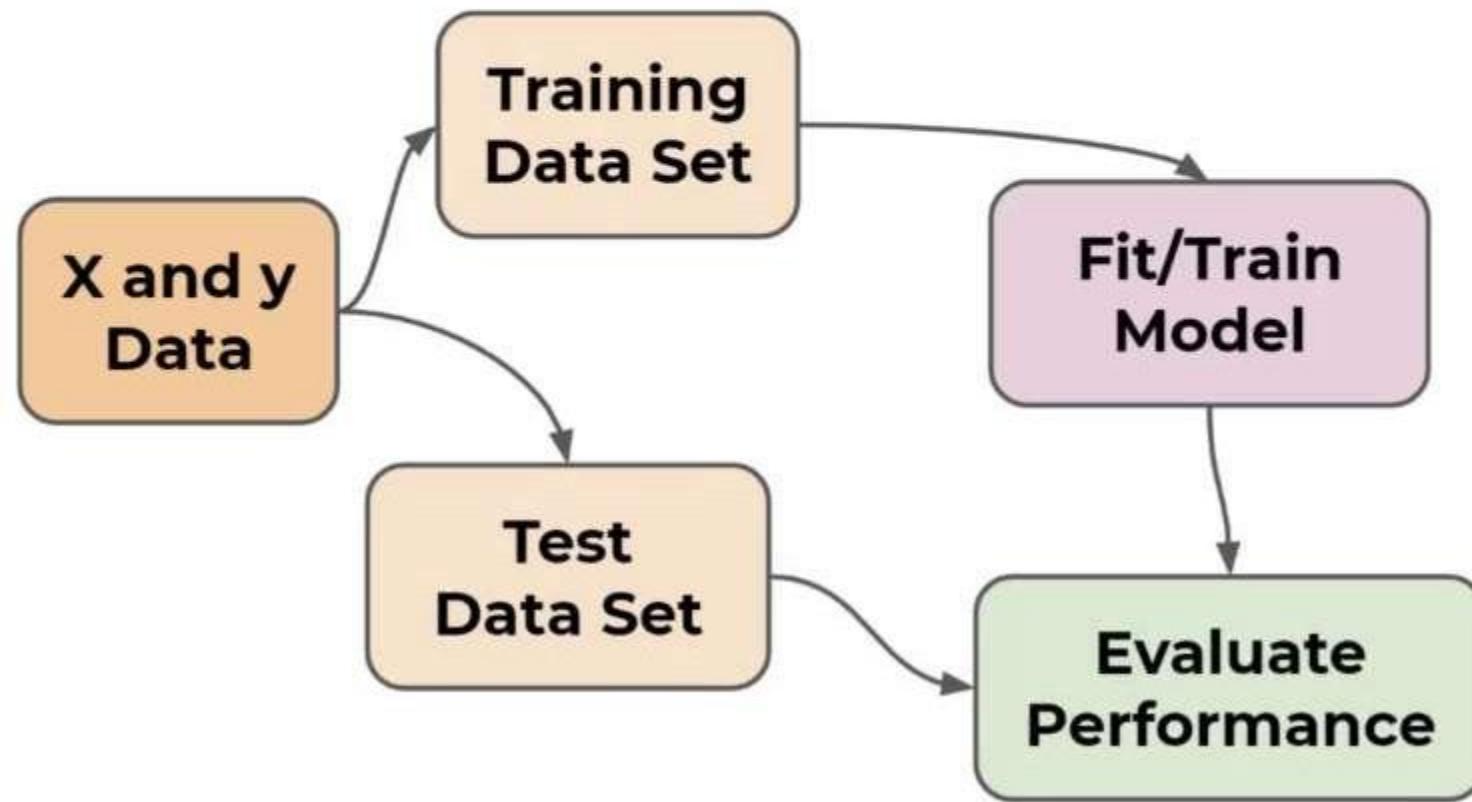
Supervised Machine Learning Process

- Fit ML Model on Training Data Set



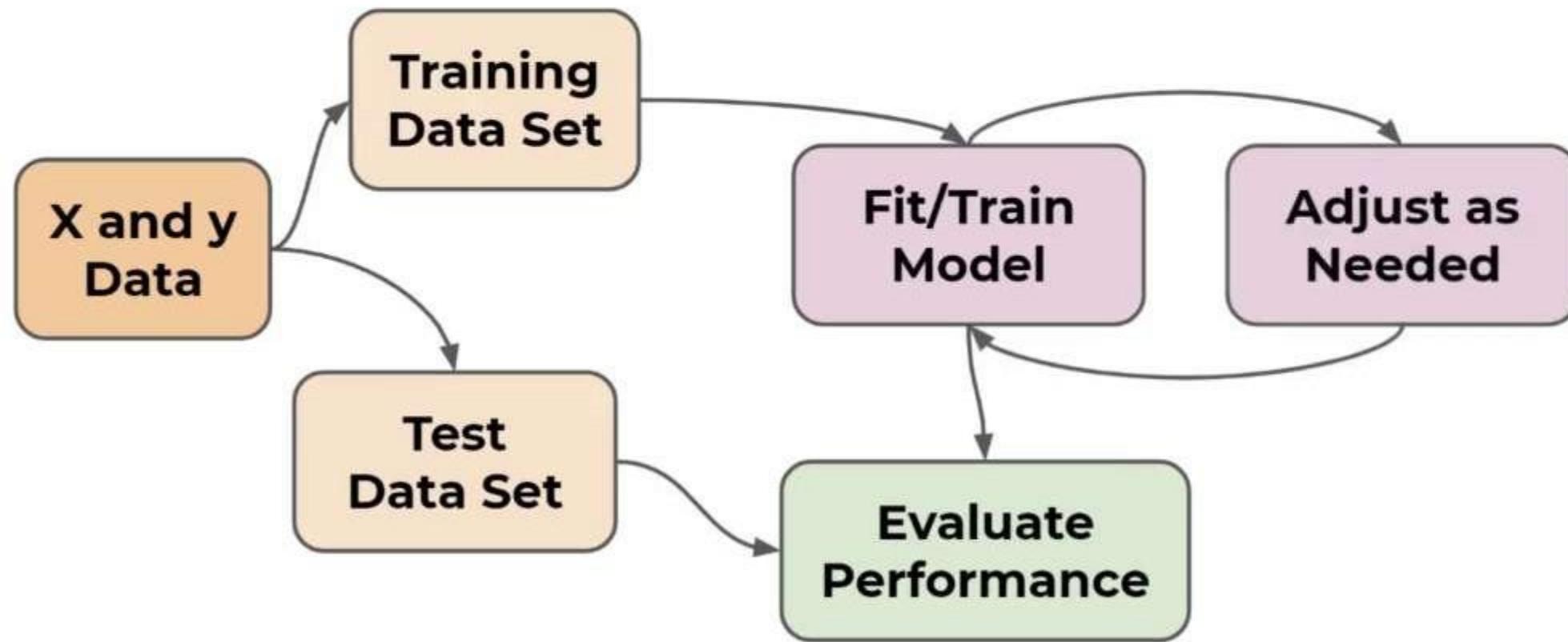
Supervised Machine Learning Process

- Evaluate Model Performance



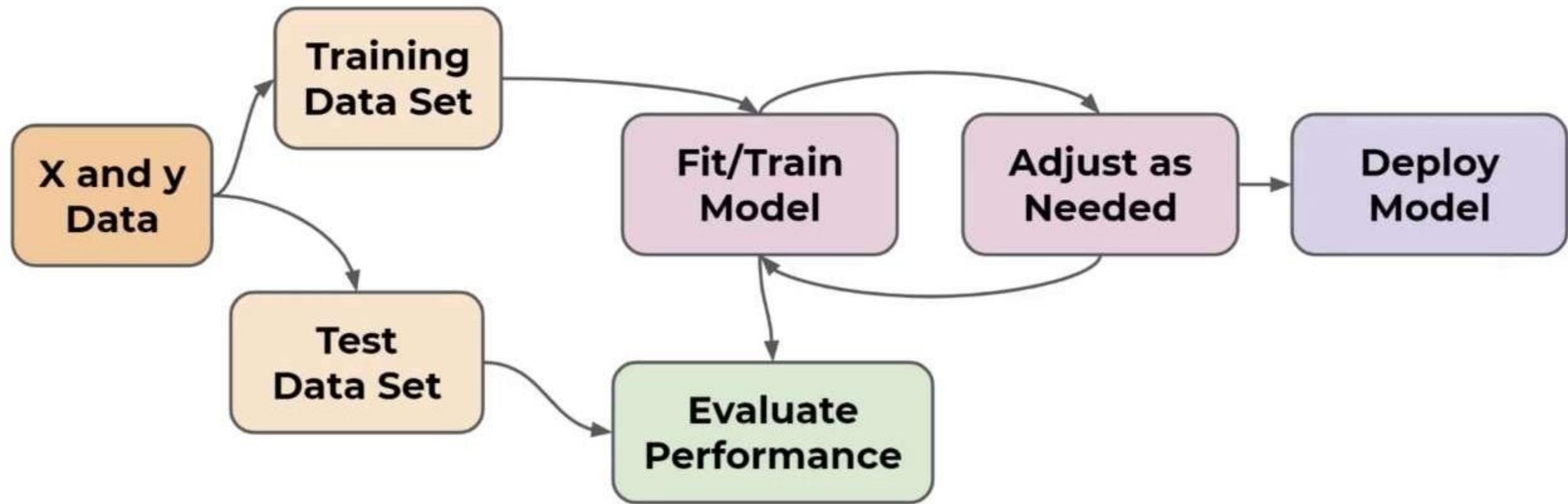
Supervised Machine Learning Process

- Adjust model hyperparameters as needed



Supervised Machine Learning Process

- Deploy model to real world



ML Pathway

