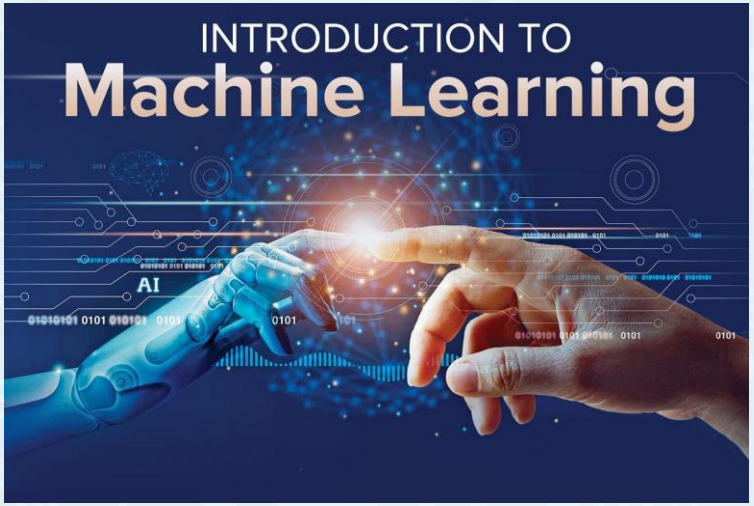
IEEE
Advancing Technology for Humanity

Introduction to machine learning (lecture 2)



By/ Aly Maher Abdel Fattah

IEEE
Galala University Student Branch

Lesson Outline

- *Introduction to the five machine learning steps*
- *Define the problem*
- *Build the dataset*
- *Model training*
- *Model evaluation*
- *Model inference*

Some Important Definitions

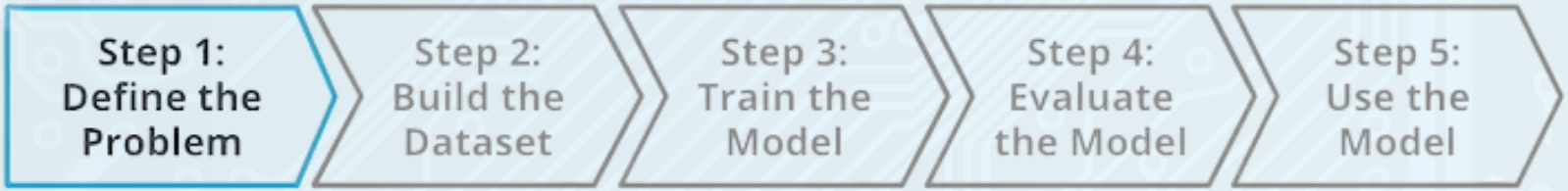
- *Clustering* ***is an unsupervised learning task that helps to determine if there are any naturally occurring groupings in the data.***
- ***A*** *categorical label* ***has a discrete set of possible values, such as "is a cat" and "is not a cat."***
- ***A*** *continuous (regression) label* ***does not have a discrete set of possible values, which means there are potentially an unlimited number of possibilities.***

Some Important Definitions

- *Discrete* ***is a term taken from statistics referring to an outcome that takes only a finite number of values (such as days of the week).***
- *A label* ***refers to data that already contains the solution.***
- ***Using*** *unlabeled data* ***means you don't need to provide the model with any kind of label or solution while the model is being trained.***

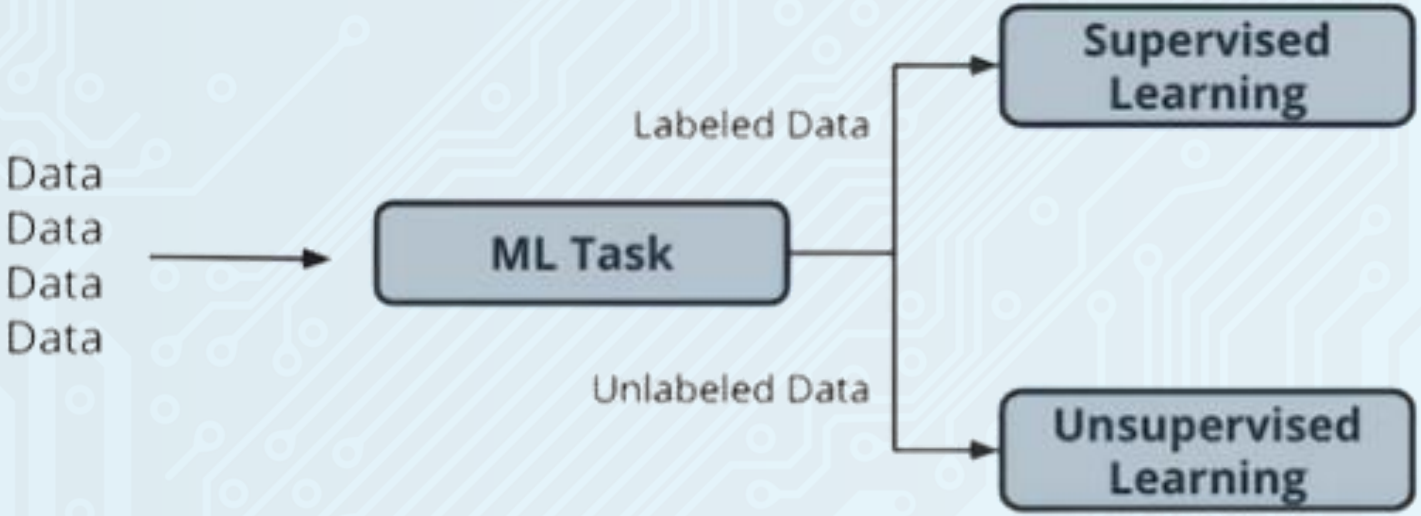
Major steps in the machine learning process

Defining a problem in machine learning

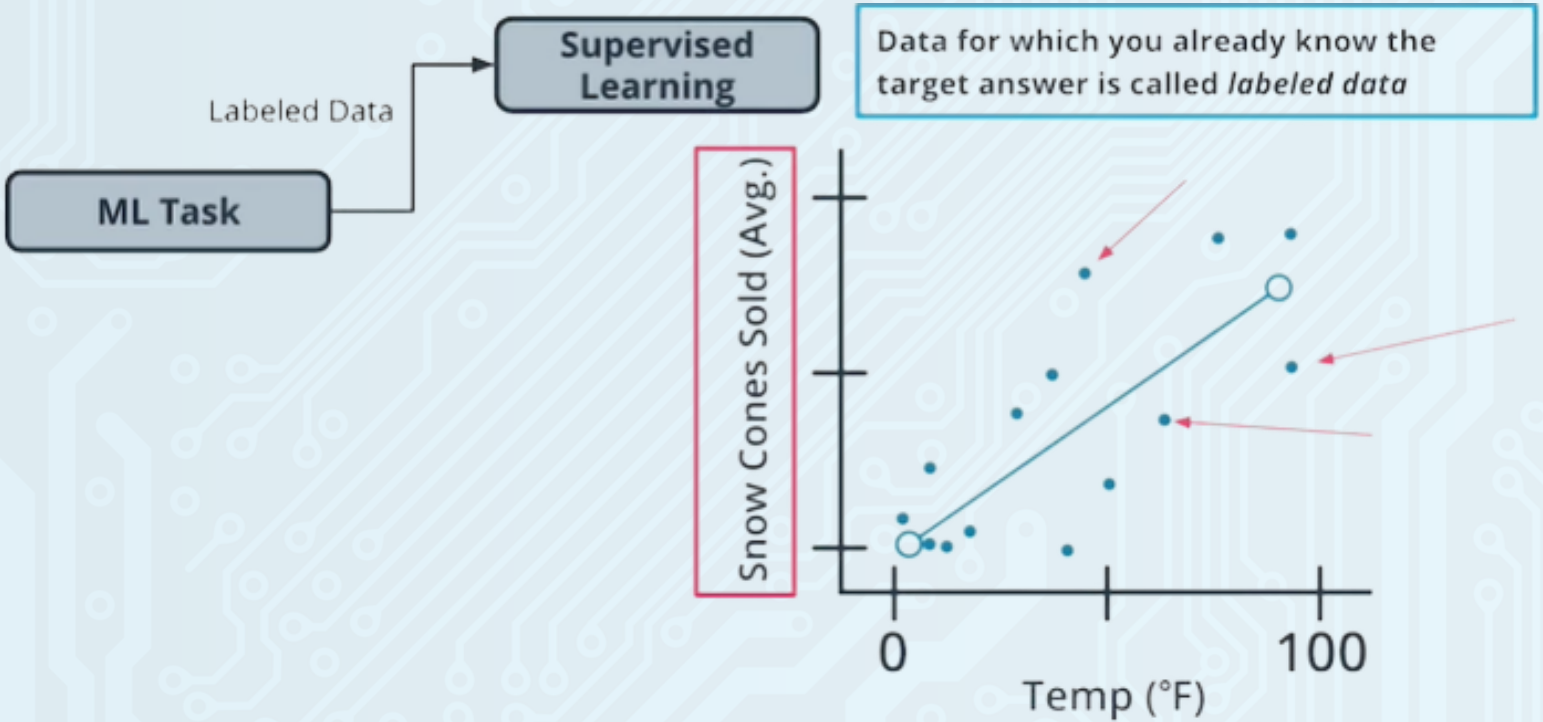


- ***Be specific!***
- ***Identify the machine learning task***

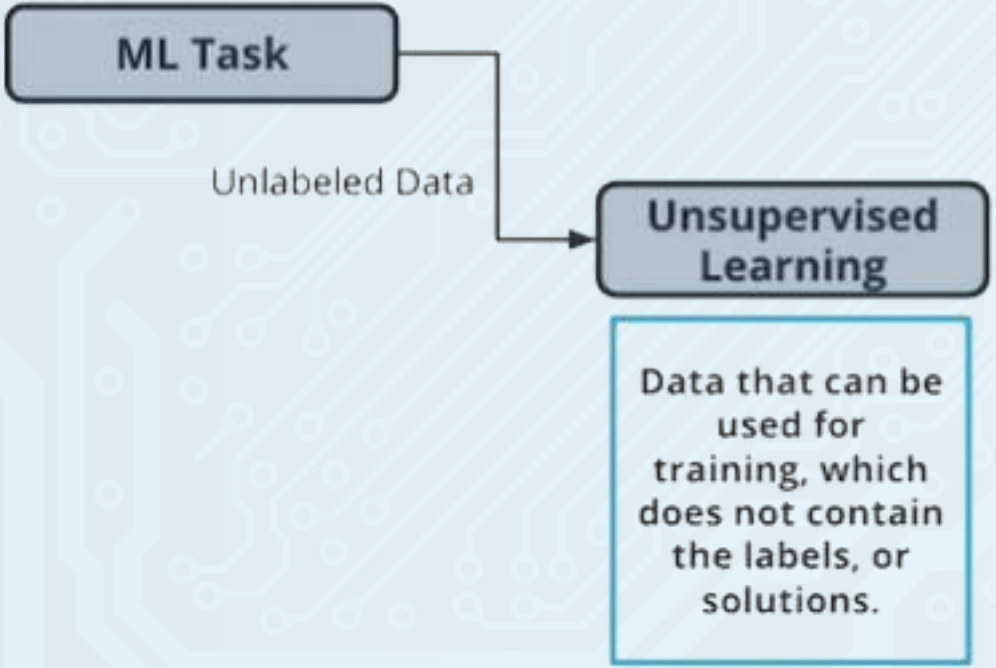
What are machine learning tasks?



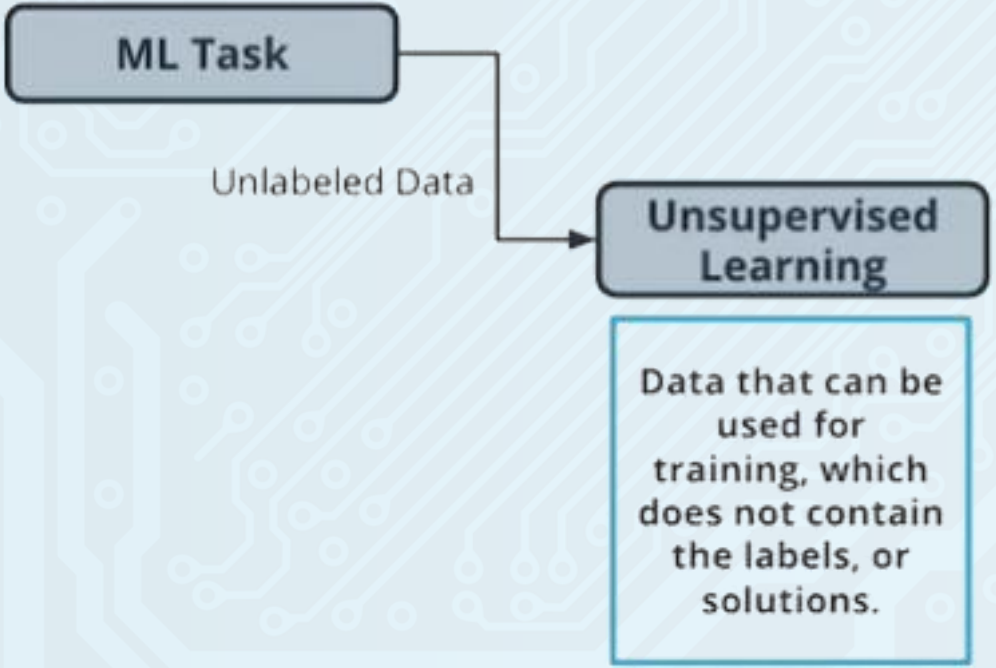
What are machine learning tasks?



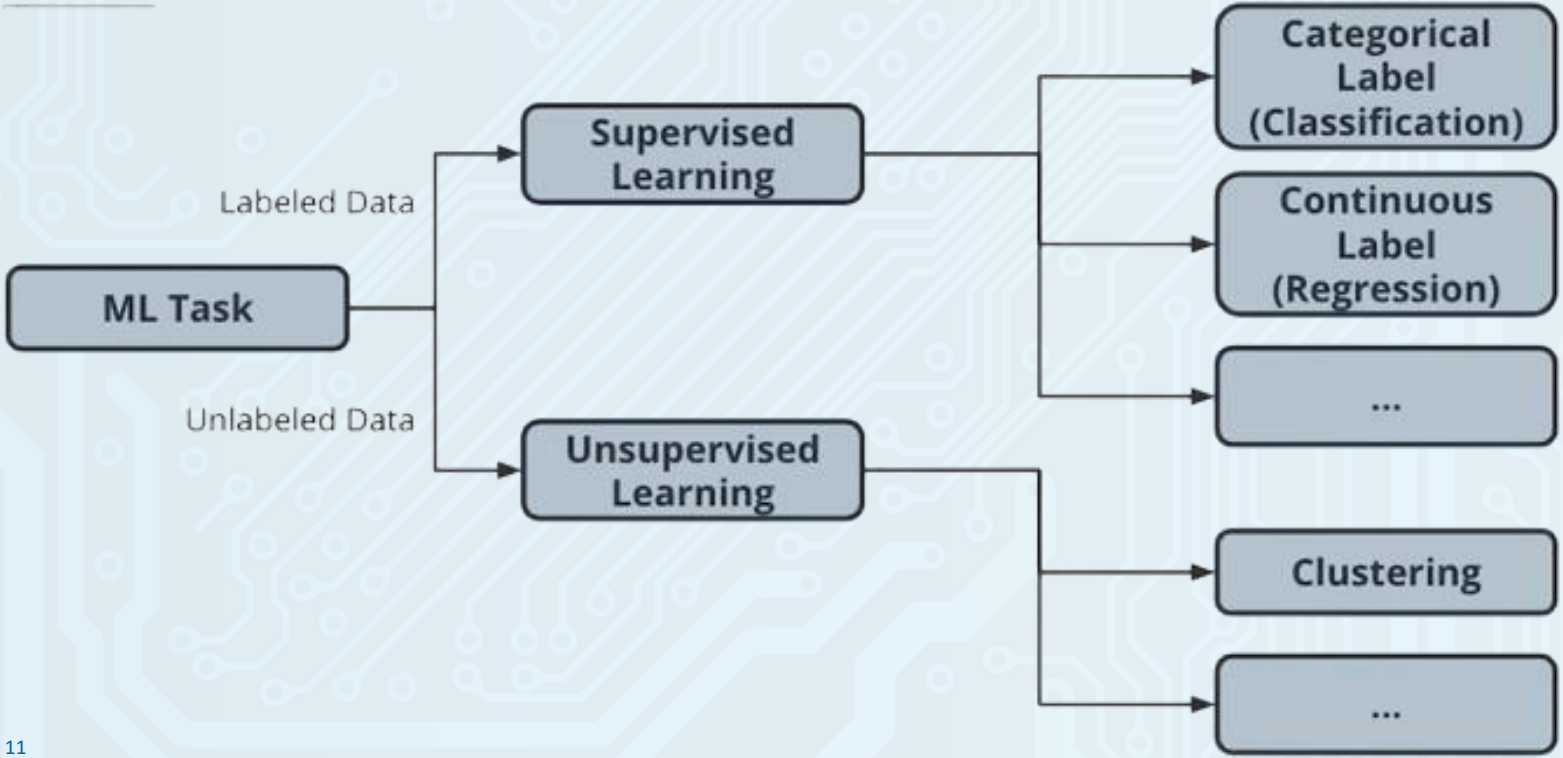
What are machine learning tasks?



What are machine learning tasks?



What are machine learning tasks?

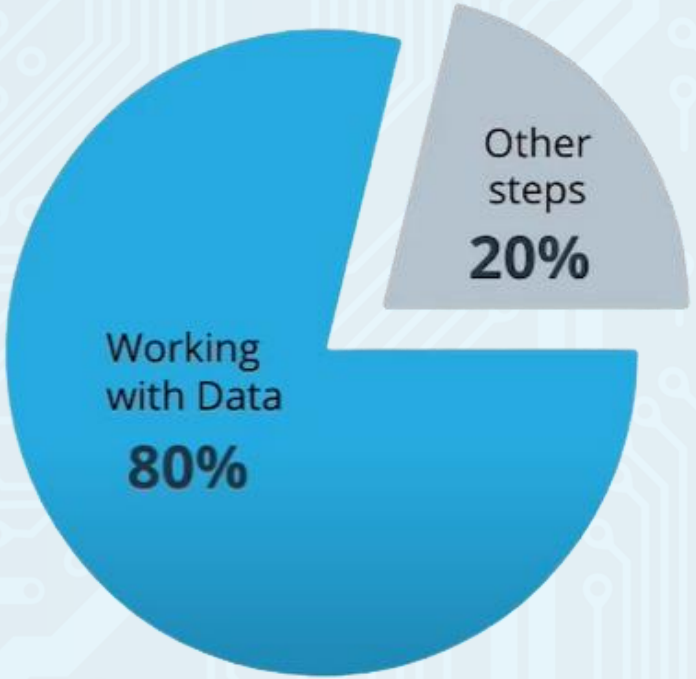


What are machine learning tasks?

Build Dataset

Step 1: Define the Problem → Step 2: Build the Dataset

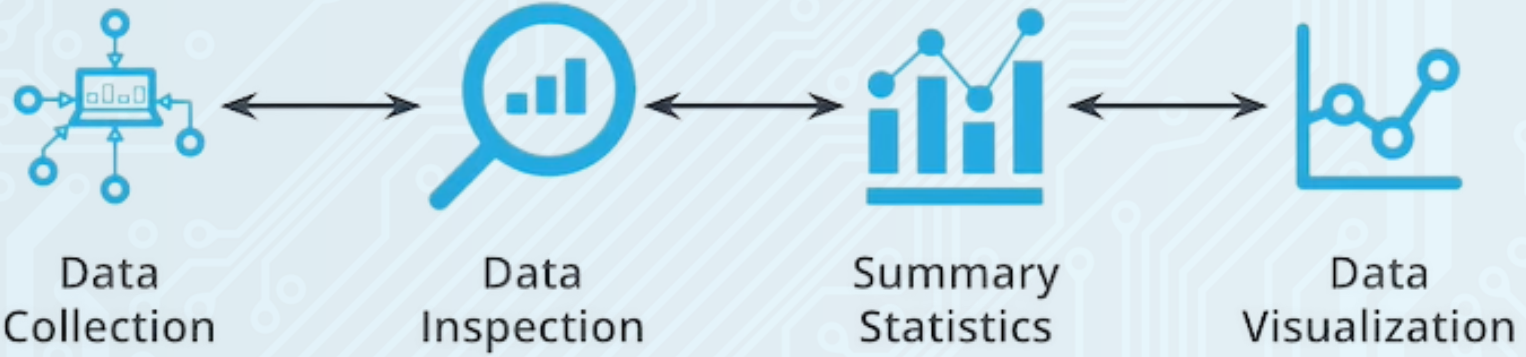
Machine learning Practitioners spend nearly 80% of their time working With data!



Time spent on machine learning project

Source: Forbes survey of 80 data scientists, March 2016

Working with data



Working with data

Data Collection

- Find and collect data related to problem you have defined
- Supervised learning ➔ Labeled Data
- Unsupervised learning ➔ Unlabeled Data

Working with data

Data Inspection

Explore your dataset looking for

- Outliers
- Missing or incomplete data
- Transform your dataset

Working with data

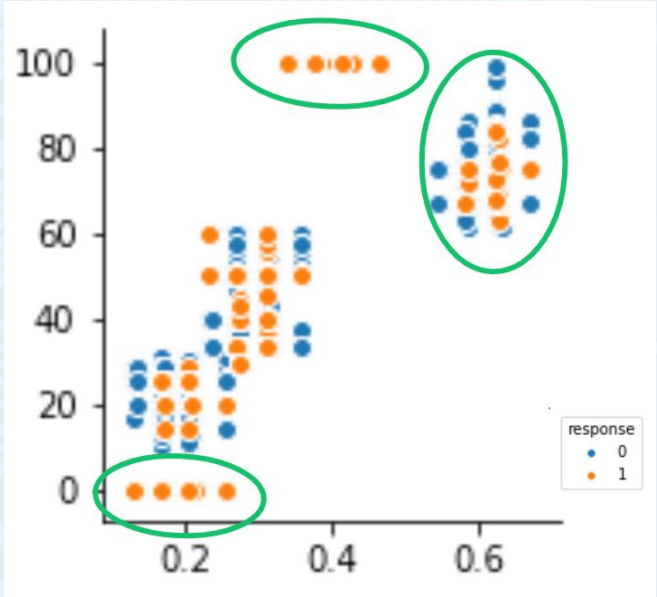
Summary Statistics

Data Visualization

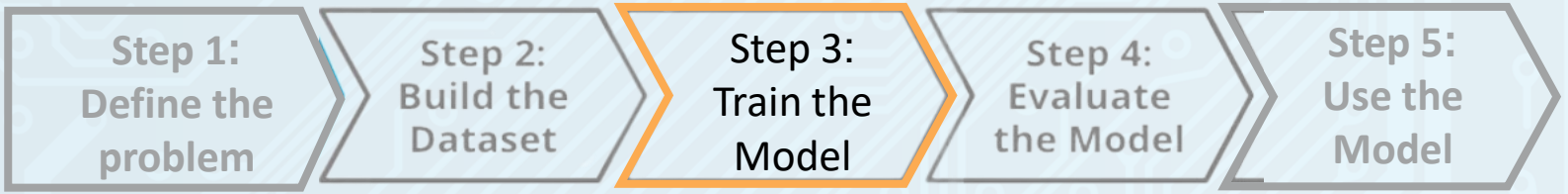
Summary statistics can Identify

- Trends in the data
- Scale of the data
- Shape of the data

Great data visualizations Communicate the findings To project stakeholders



How to Use Machine Learning



Starting a Machine Learning Task

Before you begin training you need to split your dataset

- *Majority will be held in the **training dataset***
- *The **test dataset** will be used during model evaluation*

Training a model

What does a model training algorithm actually do?

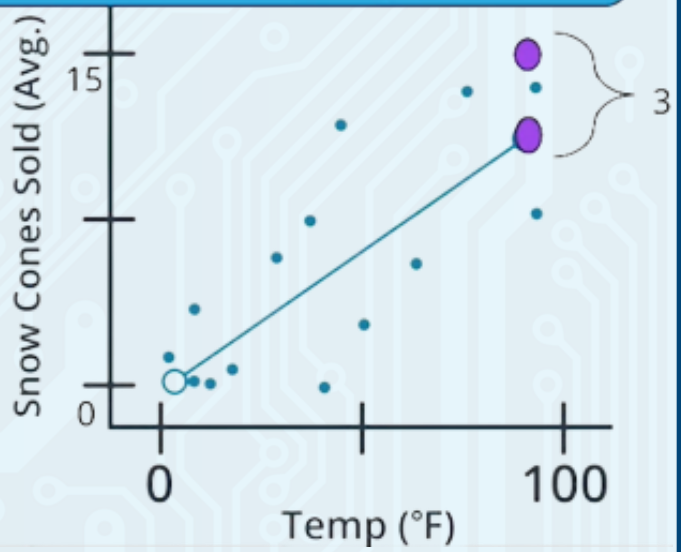
Iteratively update ***model parameters*** to minimize some ***loss function***.

- **Model Parameters**

Configuration that changes how the model behaves

- ***Loss Function***

Measurement of how close the model is to its goal



Training a model

A few other details…

- *How do I actually implement model training*
- *How do I determine which model to use*
- *Training algorithm hyperparameters*
- *Be prepared to iterate*

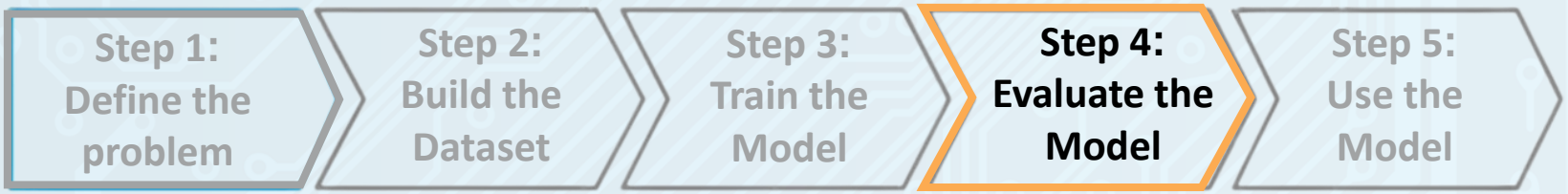
Some Important Definitions

- ***Hyperparameters** are settings on the model that are not changed during training but can affect how quickly or how reliably the model trains, such as the number of clusters the model should identify.*
- ***A loss function** is used to codify the model's distance from this goal.*

Some Important Definitions

- *Training dataset:* ***The data on which the model will be trained. Most of your data will be here.***
- *Test dataset:* ***The data withheld from the model during training, which is used to test how well your model will generalize to new data.***
- *Model parameters* ***are settings or configurations the training algorithm can update to change how the model behaves.***

How to Use Machine Learning



Evaluate the Trained Model

Model Accuracy

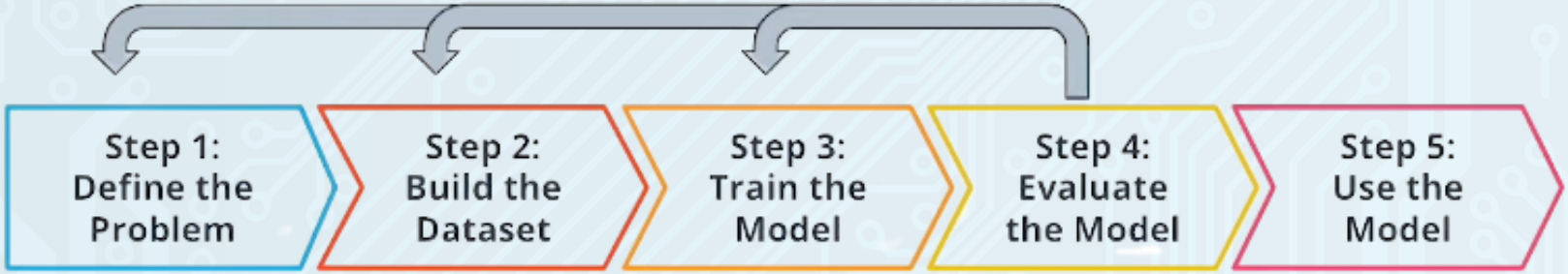
"How often your model predicts the correct species"

Evaluate the Trained Model

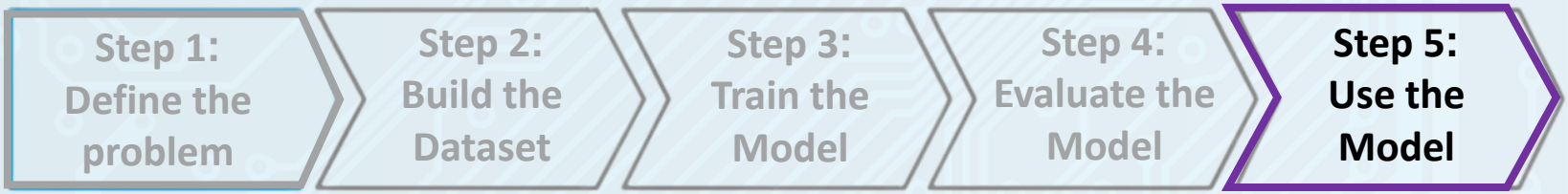
Metrics are tailored to use case

How to Use machine learning

Iterative process



How to Use Machine Learning



Inference: using your Model

- ***Use your model to solve real problems***
- ***Monitor the results***

IEEE
Advancing Technology for Humanity

Thank you

Any Question?

IEEE
Galala University Student Branch