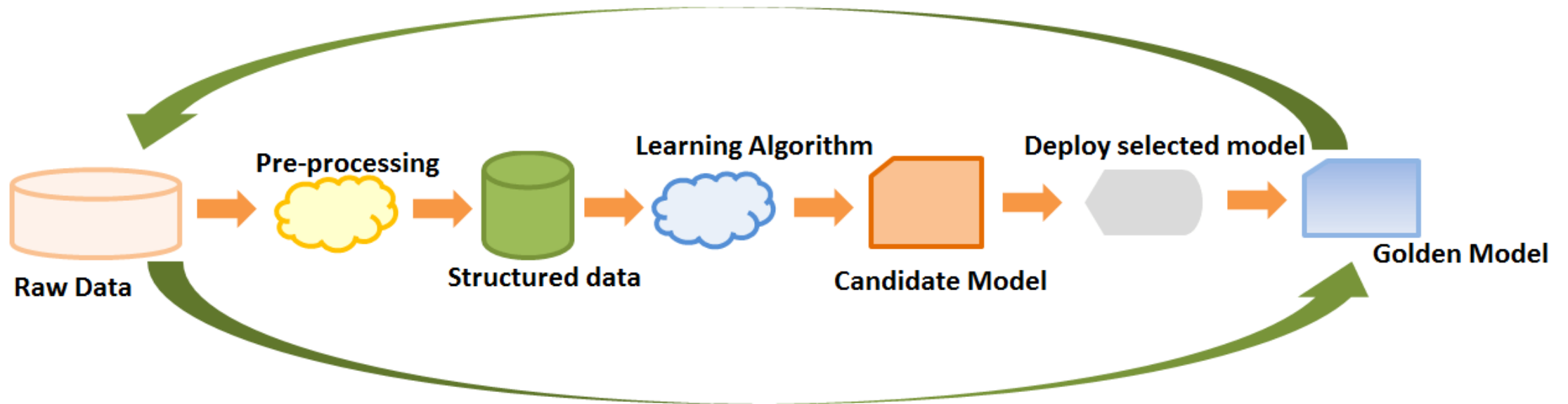


Machine Learning Process

Machine learning is **the process of making systems that learn and improve by themselves, by being specifically programmed.**

The ultimate goal of machine learning is to design algorithms that automatically help a system gather data and use that data to learn more.



Steps of Machine Learning(7)

Data Collection

Data Preparation

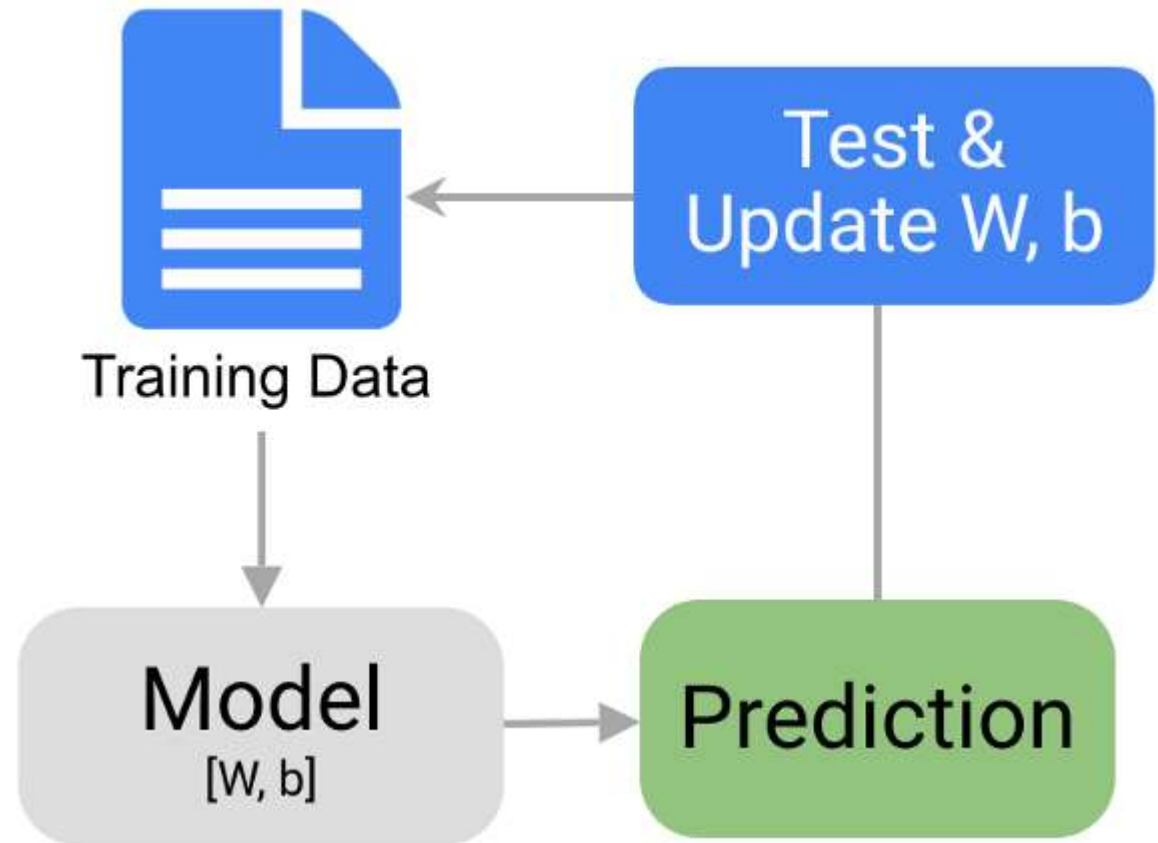
Choose a Model

Train the Model

Evaluate the Model

Parameter Tuning

Make Predictions



Data Collection

The quantity & quality of your data dictate how accurate our model is

The outcome of this step is generally a representation of data (simplifies to specifying a table) which we will use for training

Using pre-collected data, by way of datasets from Kaggle, UCI, etc., still fits into this step.

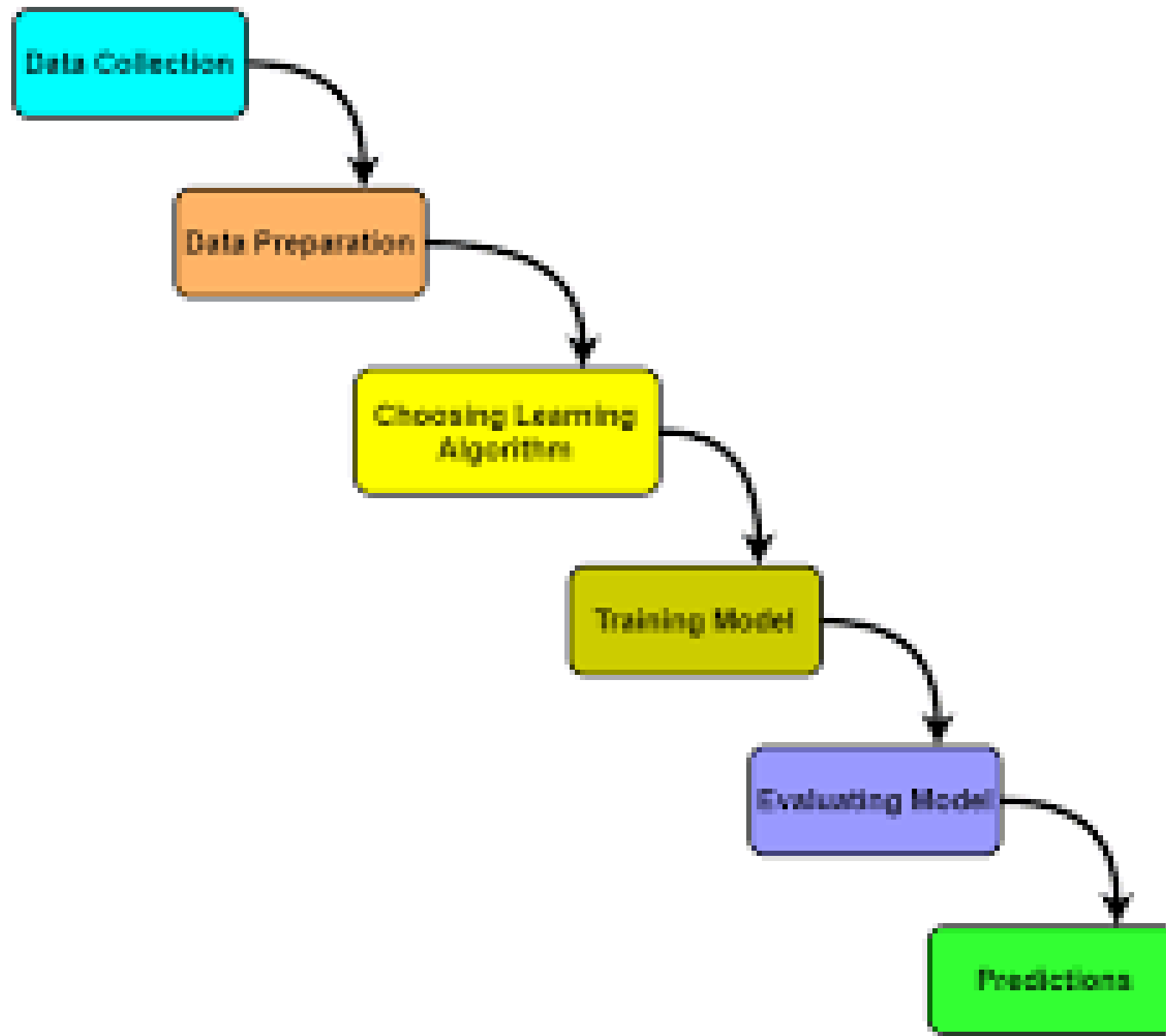
Data Preparation

Wrangle data and prepare it for training

Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.)

Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data

Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis and Split into training and evaluation sets



Machine Learning Workflow

Choose a Model

Different algorithms are for different tasks; choose the right one

4 - Train the Model

The goal of training is to answer a question or make a prediction correctly as often as possible

Linear regression example: algorithm would need to learn values for m (or W) and b (x is input, y is output)

Each iteration of process is a training step.

Evaluate the Model

Uses some metric or combination of metrics to "measure" objective performance of model

Test the model against previously unseen data

This unseen data is meant to be somewhat representative of model performance in the real world, but still helps tune the model.

Good train/eval split? 80/20, 70/30, or similar, depending on domain, data availability, dataset particulars, etc.

Parameter Tuning

This step refers to *hyperparameter* tuning, which is an "artform" as opposed to a science

Tune model parameters for improved performance

Simple model hyperparameters may include: number of training steps, learning rate, initialization values and distribution, etc.

Make Predictions

Using further (test set) data which have, until this point, been withheld from the model are used to test the model; a better approximation of how the model will perform in the real world

Workflow of Machine Learning

A blueprint for solving machine learning problems.

1. Defining the problem and assembling a dataset
2. Choosing a measure of success
3. Deciding on an evaluation protocol
4. Preparing your data
5. Developing a model that does better than a baseline
6. Scaling up: developing a model that overfits
7. Regularizing your model and tuning your parameters

