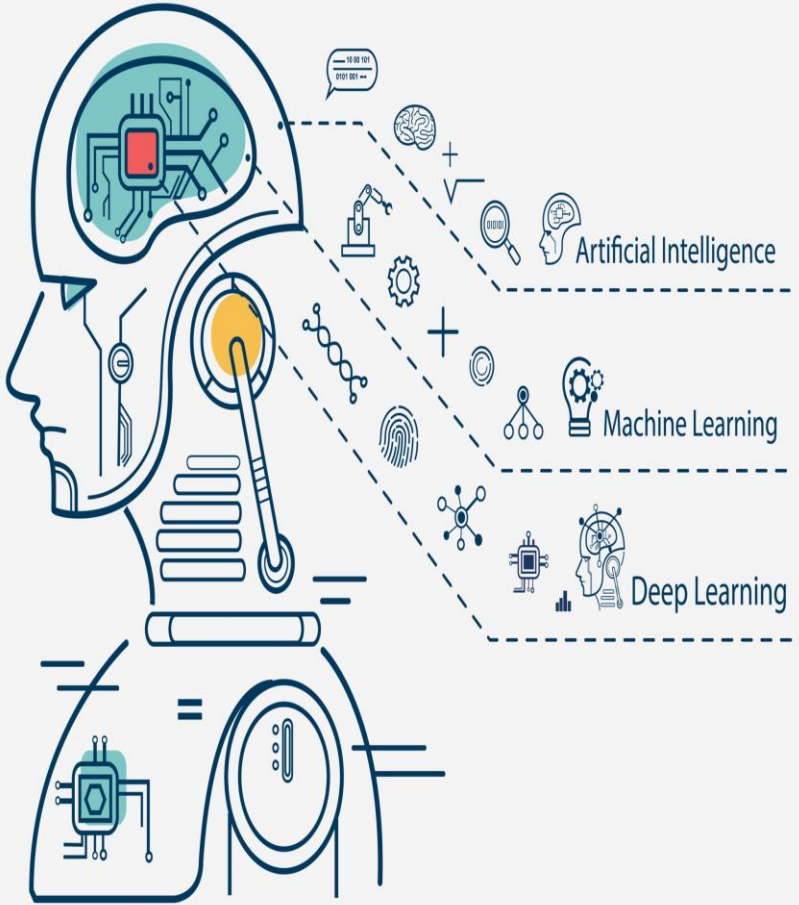
Machine Learning

(Introduction)



Machine Learning

Topics to be covered

- What is Machine Learning?
- What is Deep Learning?
- Supervised and Unsupervised Learning
- Supervised Learning Process
- Types of Data Sets
- Reinforcement Learning

Companion Book

- We will be using **Introduction to Statistical Learning** by Gareth James as a companion book.
- Read Chapter 1and 2.
- Students who want the mathematical theory should do the reading.
- Students who just want light theory and more interested in Python Applications.

What is Machine Learning?

- Machine learning is a method of data analysis that automates analytical model building.
- Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

What is it used for?

- Fraud detection.
- Web search results.
- Real-time ads on web pages
- Credit scoring.
- Prediction of equipment failures.
- New pricing models.
- Network intrusion detection.
- Recommendation Engines
- Customer Segmentation
- Text Sentiment Analysis
- Customer Churn
- Pattern and image recognition.
- Email spam filtering.

Machine Learning

- There are different types of machine learning:
  - Supervised Learning
  - Unsupervised Learning

Machine Learning

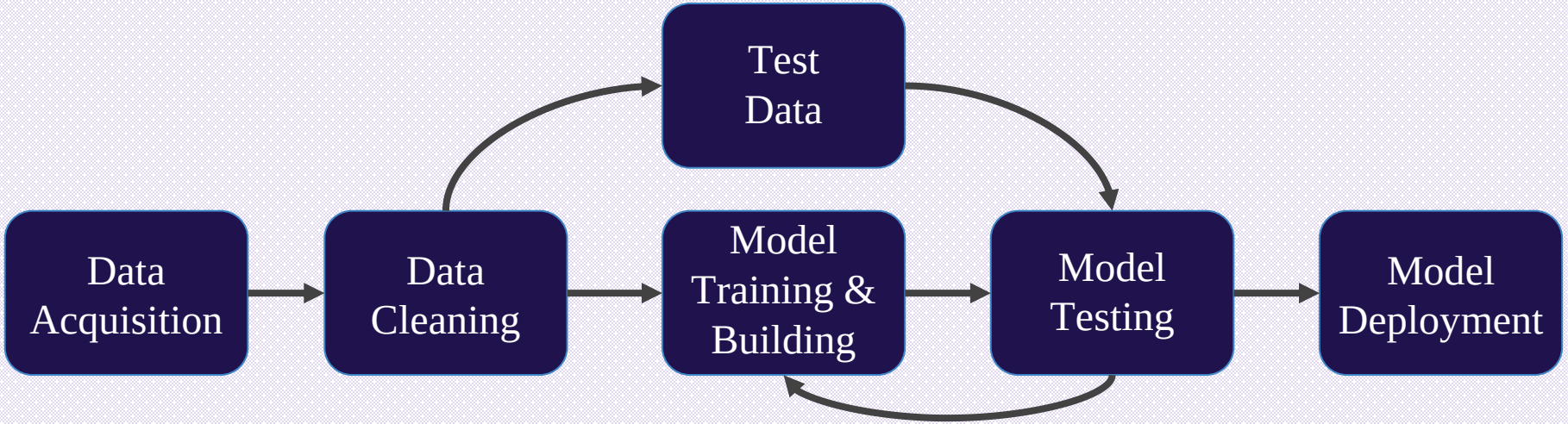
- **Machine Learning**
  - Automated analytical models.
- **Neural Networks**
  - A type of machine learning architecture modeled after biological neurons.
- **Deep Learning**
  - A neural network with more than one hidden layer.

Read these topics below:

Difference between Machine Learning and Deep Learning
https://www.educative.io/blog/deep-vs-machine-learning

Machine Learning Vs Deep Learning Vs AI
https://wiki.pathmind.com/ai-vs-machine-learning-vs-deep-learning

Machine Learning Process



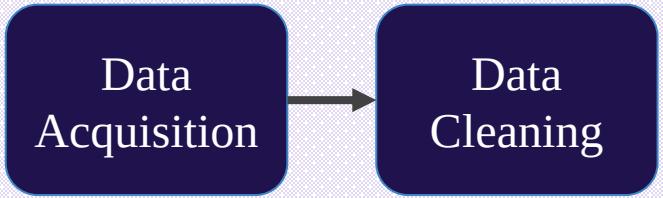
Machine Learning Process

- Get your data! Customers, Sensors, etc...

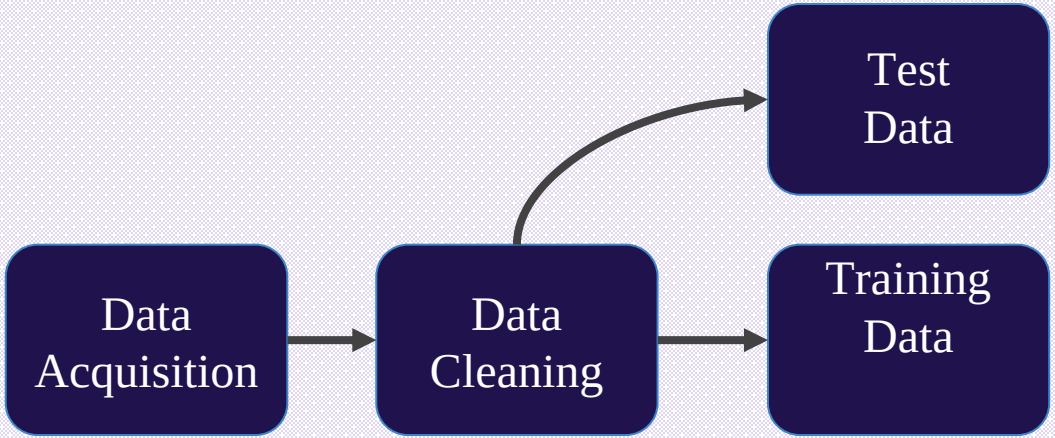
Data Acquisition

Machine Learning Process

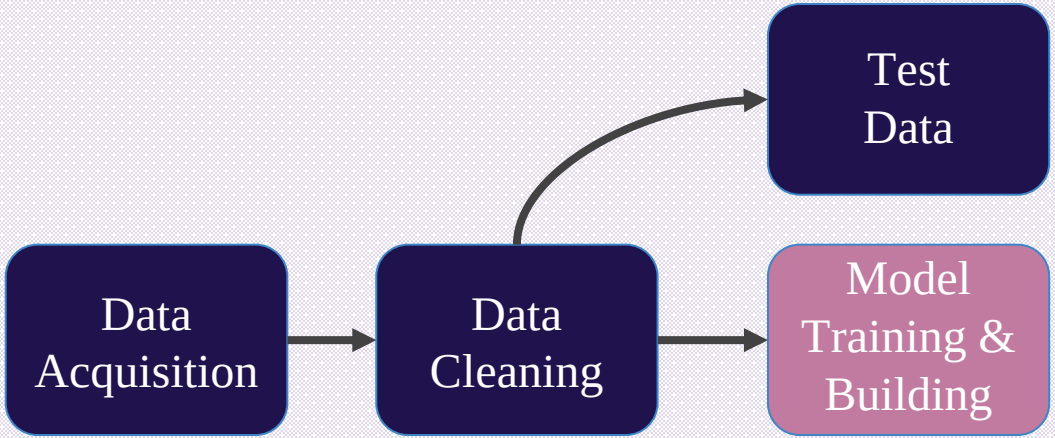
- Clean and format your data (using Pandas)



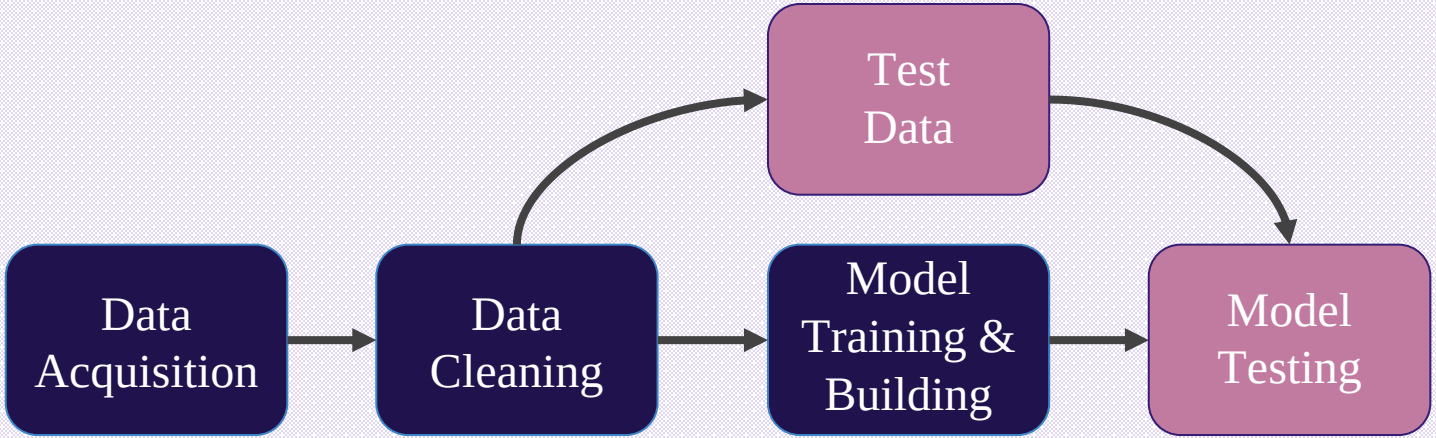
Machine Learning Process



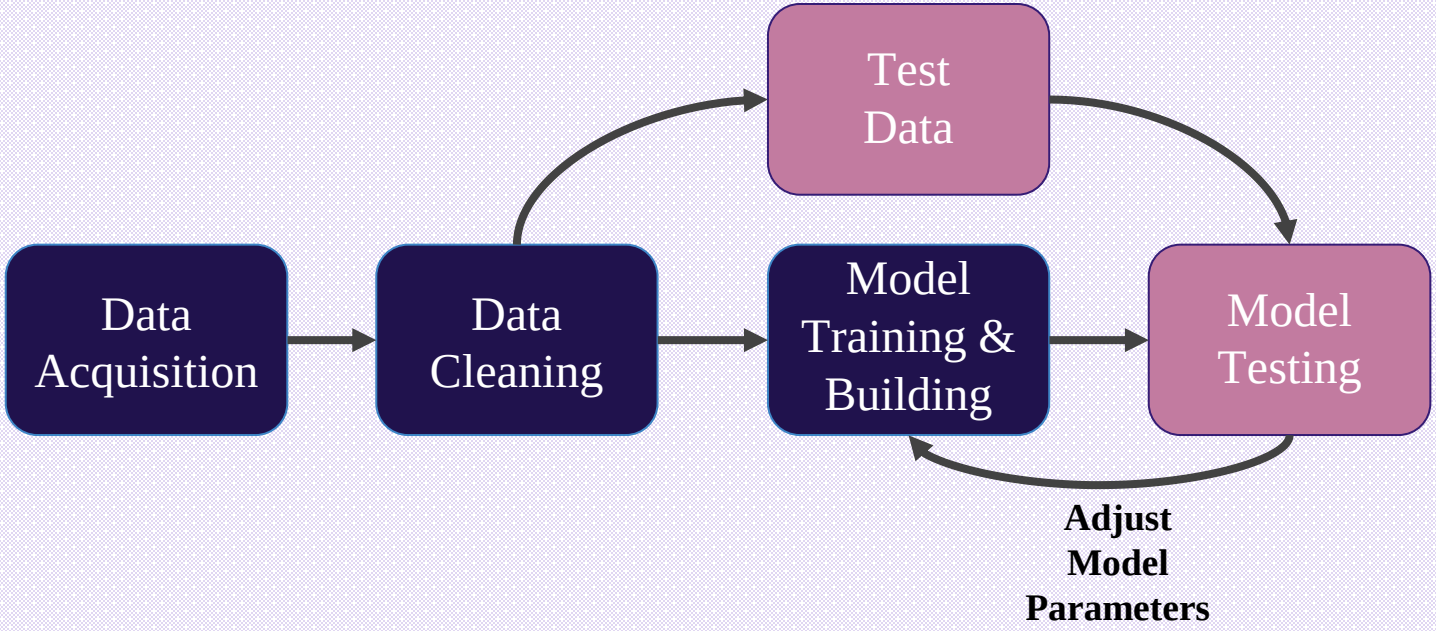
Machine Learning Process



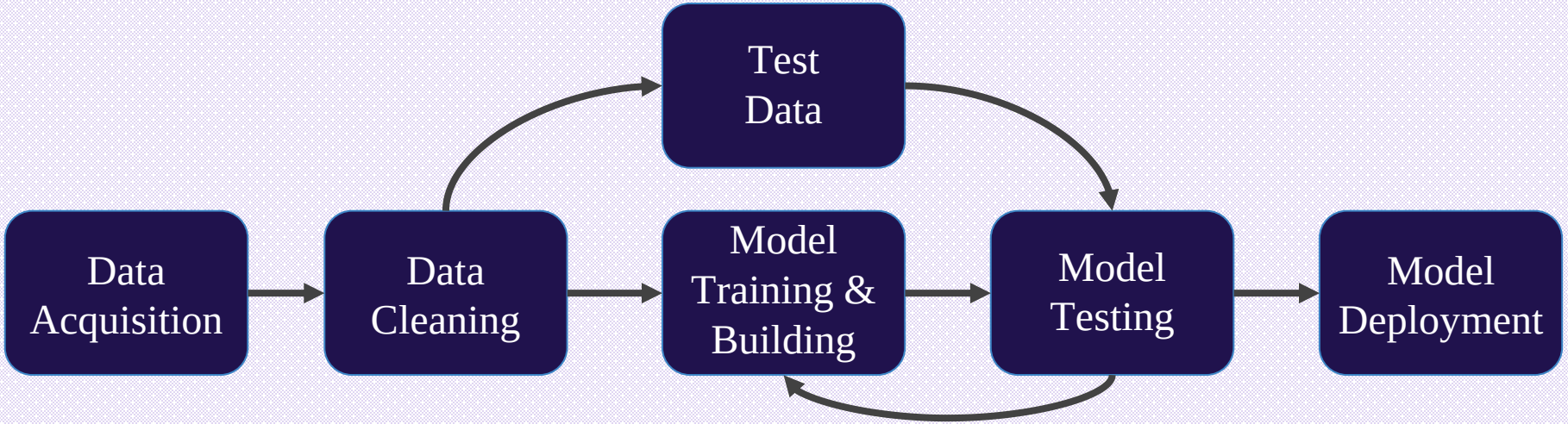
Machine Learning Process



Machine Learning Process



Machine Learning Process



1. Supervised Learning

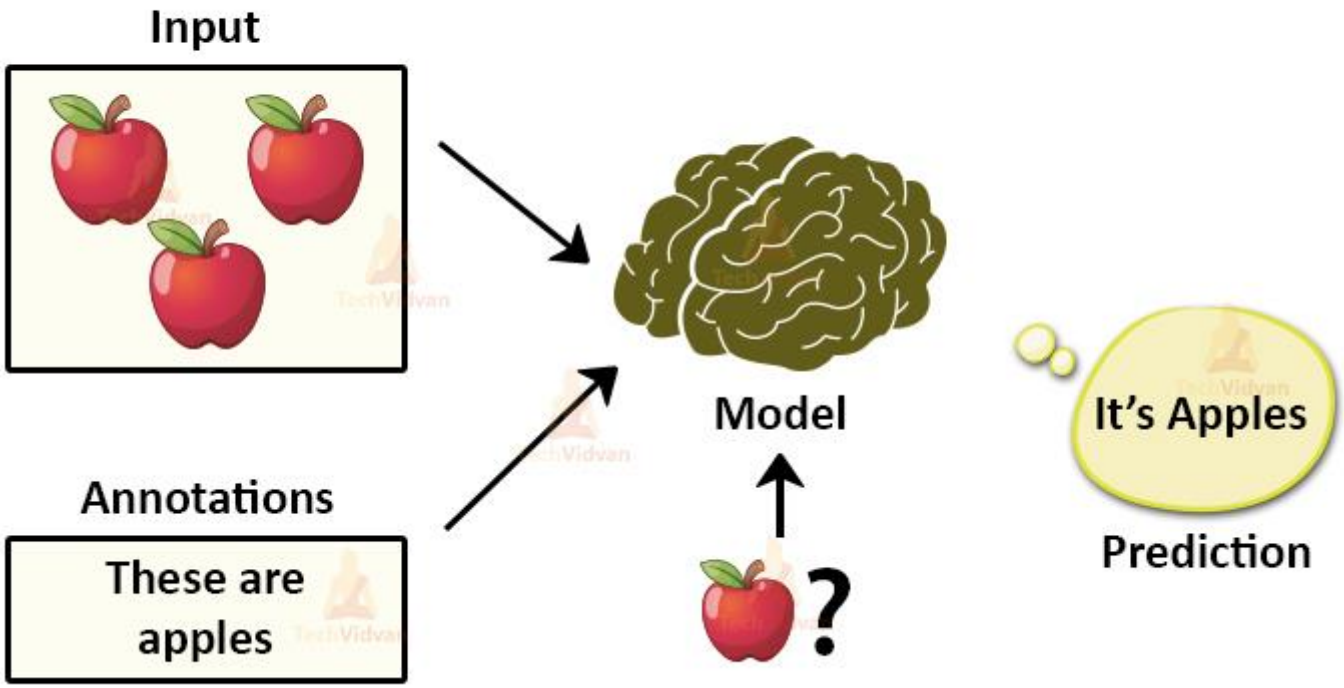
Supervised Learning

- **Supervised learning** algorithms are trained using **labeled** examples, such as an input where the desired output is known.
- Supervised learning is commonly used in applications where historical data predicts likely future events.
- For example, a piece of equipment could have data points labeled either “F” (failed) or “R” (runs).
- For example, a segment of text could have a category label, such as:
  - **Spam** vs. **Legitimate** Email
  - **Positive** vs. **Negative** Movie Review

Supervised Learning

- For example, it can anticipate when credit card transactions are likely to be fraudulent or which insurance customer is likely to file a claim.
- Or it can attempt to predict the price of a house based on different features for houses for which we have historical price data.

Supervised Learning in ML



Supervised Learning

- The learning algorithm receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors.
- It then modifies the model accordingly.

Supervised Learning

- Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabeled data.

Supervised Learning

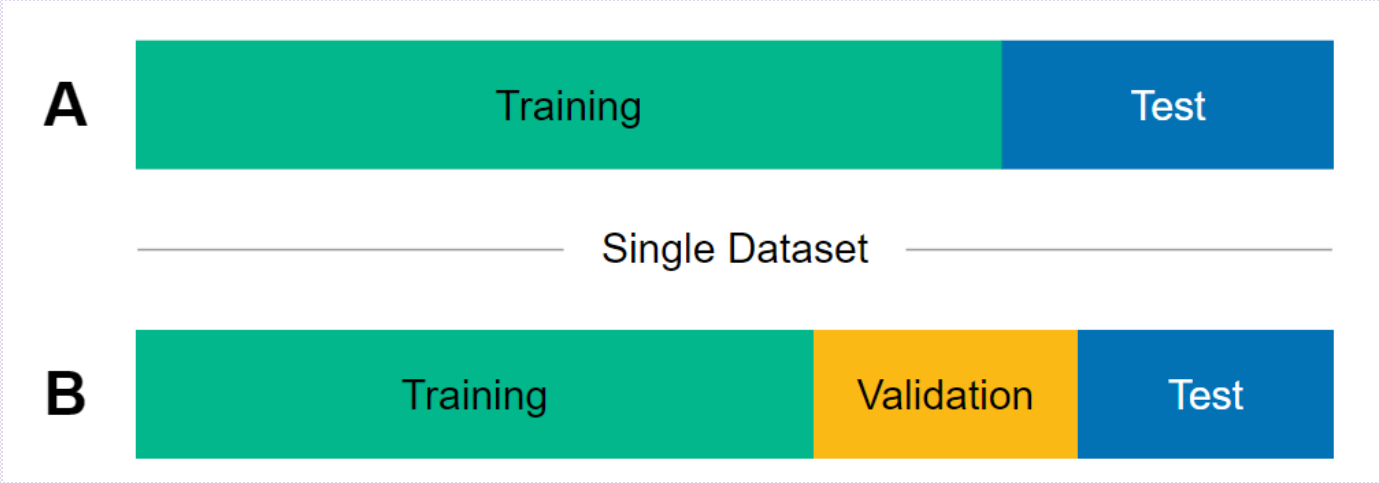
- The network receives a set of inputs along with the corresponding correct outputs, and the algorithm learns by comparing its actual output with correct outputs to find errors.
- It then modifies the model accordingly.

Supervised Learning

- What we just showed is a simplified approach to supervised learning, it contains an issue!
- Is it fair to use our single split of the data to evaluate our models performance?
- After all, we were given the chance to update the model parameters again and again.

Need for Validation data?

https://developers.google.com/machine-learning/crash-course/validation/another-partition



Supervised Learning

- To fix this issue, data is often split into **3 sets**
  - **Training Data**
    - Used to train model parameters
  - **Validation Data**
    - Used to determine what model hyperparameters to adjust
  - **Test Data**
    - Used to get some final performance metric

Supervised Learning

- This means after we see the results on the **final test set** we don’t get to go back and adjust any model parameters!
- This final measure is what we label the true performance of the model to be.

Supervised Learning

- In general we will simplify our data by using a simple **train/test split.**
- We will simply train and then evaluate on a test set .
- Later on, you will be able to easily perform another split to get **3 data sets** if you desire.

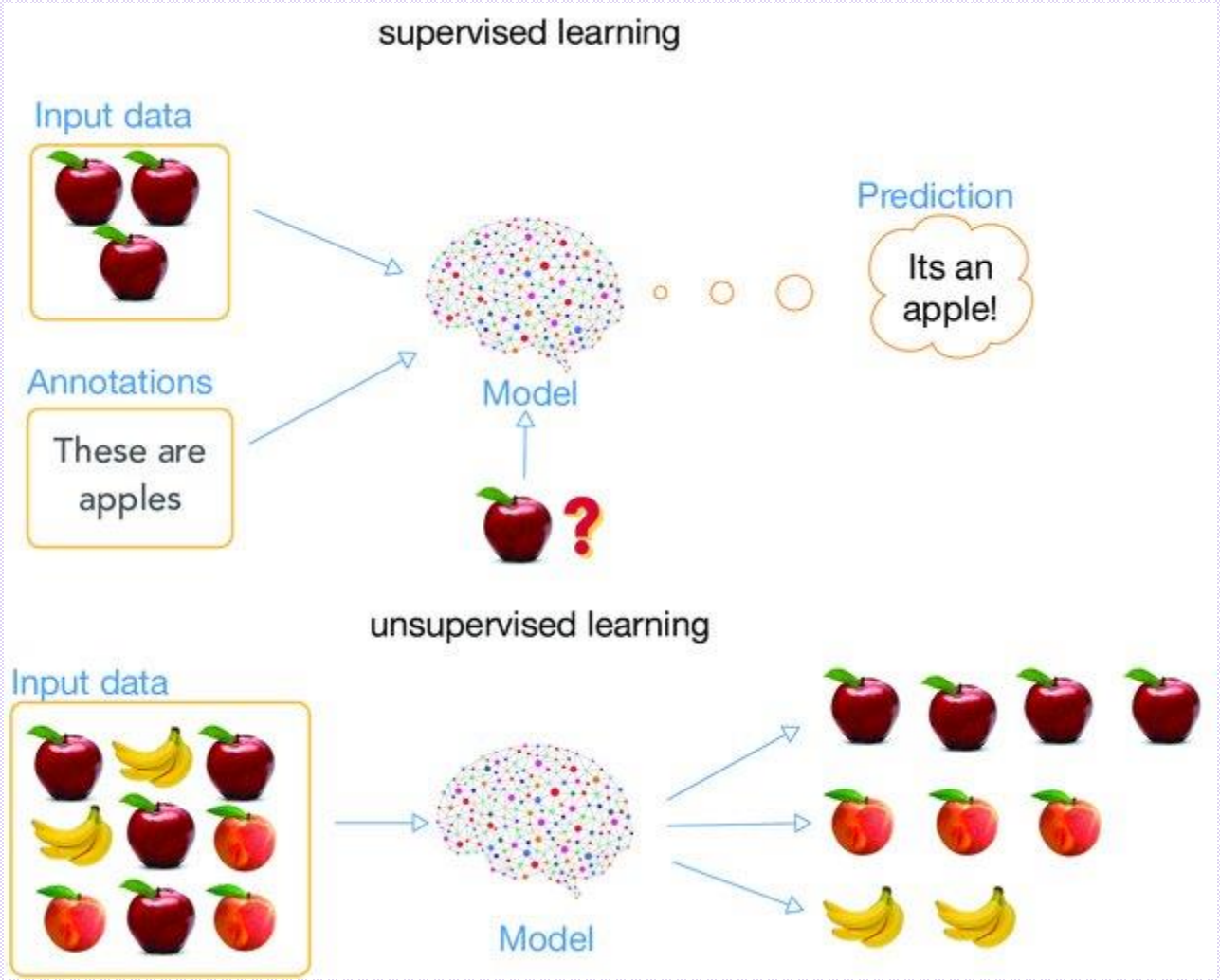
2. Unsupervised Learning

Machine Learning

- We’ve covered supervised learning, where the **label was known** due to **historical labeled data.**
- But what happens when we don’t have historical labels?

Unsupervised Learning

- **Unsupervised learning** is used against data that has no historical labels.
- The system is not told the "right answer." The algorithm must figure out what is being shown.
- The goal is to explore the data and find some structure within.


supervised learning
Input data
Prediction
Its an apple!
Annotations
These are apples
Model
unsupervised learning
Input data
Model

Unsupervised Learning

- Or it can find the main attributes that separate customer segments from each other.
- Popular techniques include self-organizing maps, nearest-neighbor mapping, k-means clustering and singular value decomposition.

Unsupervised Learning

- These algorithms are also used to segment text topics, recommend items and identify data outliers.

Unsupervised Learning

- There are certain tasks that fall under unsupervised learning:
  - Clustering
  - Anomaly Detection
  - Dimensionality Reduction

Unsupervised Learning

- **Clustering**
  - Grouping together **unlabeled** data points into categories/clusters
  - Data points are assigned to a cluster based on similarity

Unsupervised Learning

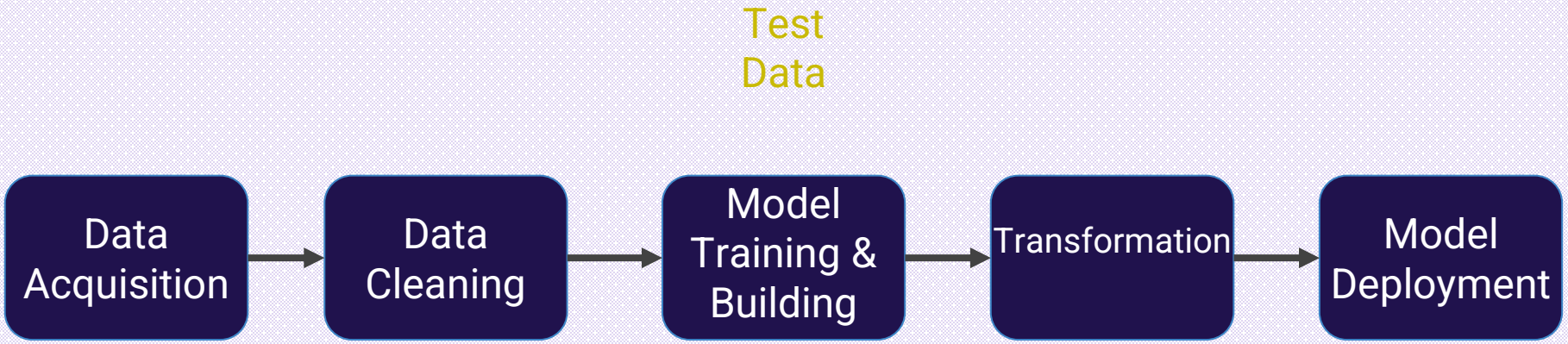
- **Anomaly Detection**
  - Attempts to detect outliers in a dataset
  - For example, fraudulent transactions on a credit card.

Unsupervised Learning

- **Dimensionality Reduction**
  - Data processing techniques that reduces the number of features in a data set, either for compression, or to better understand underlying trends within a data set.

- **Unsupervised Learning**
  - It’s important to note, these are situations where we **don’t** have the correct answer for historical data!
  - Which means evaluation is much harder and more nuanced!

Unsupervised Process



3. Reinforcement Learning

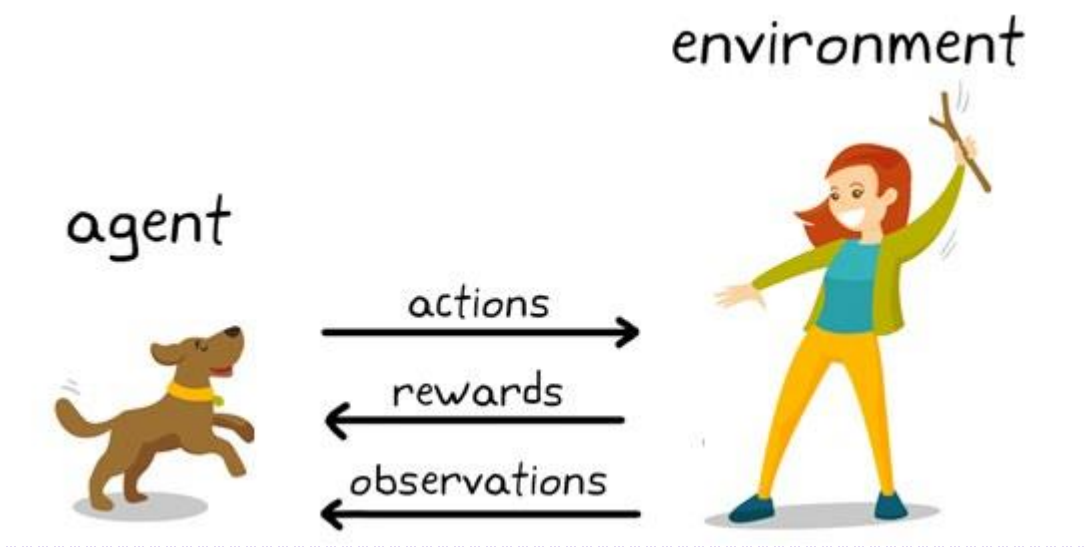
Reinforcement Learning

- **Reinforcement learning** is often used for robotics, gaming and navigation.
- With reinforcement learning, the algorithm discovers through trial and error which actions yield the greatest rewards.

Reinforcement Learning

- This type of learning has three primary components: the **agent** (the learner or decision maker), **the environment** (everything the agent interacts with) and **actions** (what the agent can do).

Example



https://www.kdnuggets.com/2019/10/mathworks-reinforcement-learning.html

Reinforcement Learning

- The objective is for the agent to choose actions that maximize the expected reward over a given amount of time.
- The agent will reach the goal much faster by following a good policy.

Reinforcement Learning

- So the goal in reinforcement learning is to learn the best policy.