Unsupervised Learning

Unsupervised Learning

- It is now time to begin learn about machine learning algorithms used for Unsupervised Learning!
- This will be a paradigm shift from our previous discussions on Supervised Learning.

Unsupervised Learning

- If Data Science is a mix between an art and a mathematical science, unsupervised learning is where we get to dive deeper into the art.

Unsupervised Learning

- Supervised Learning
  - Using historical **labeled** data, predict a label on new data (regression or classification).
- Unsupervised Learning
  - Using **unlabeled** data, discover patterns, clusters, or significant components.

Unsupervised Learning

- Unsupervised Learning:
  - Clustering:
    - Using features, group together data rows into distinct clusters.
  - Dimensionality Reduction:
    - Using features, discover how to combine and reduce into fewer components.

Unsupervised Learning

- Paradigm shift for supervised to unsupervised learning:
  - ***Supervised performance metrics will not apply for unsupervised learning!***
  - How can we compare to a correct label answer, if there was no label to begin with?

Unsupervised Learning

- Instead of metrics like RMSE or Accuracy, we will need to figure out other ways of assessing unsupervised model performance or reasonableness.
- Even our understanding of what “performance” actually means will need to change with unsupervised learning!

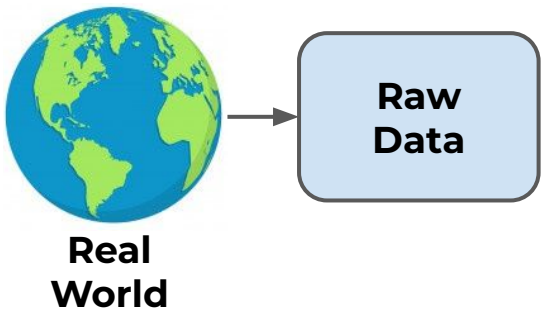
Unsupervised Learning

- What does our Machine Learning Pathway look like with Unsupervised Learning?

Unsupervised ML Pathway

Real World

ML Pathway



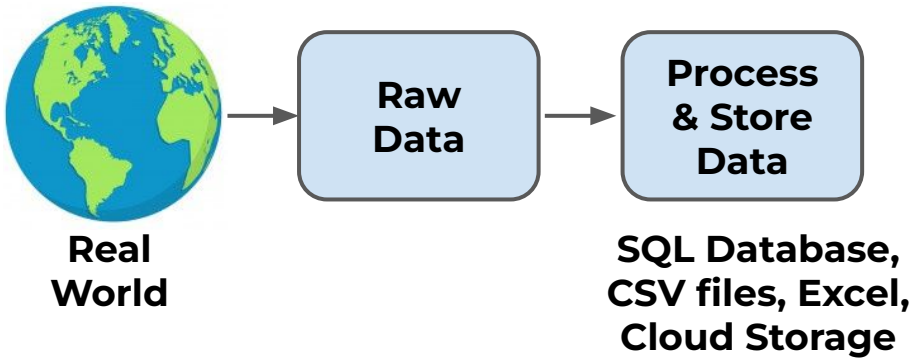
ML Pathway

Real World

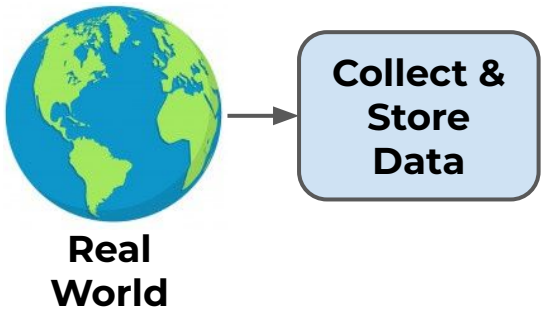
Raw Data

Physical Sensors, Surveys, Simulations, Experiments, Data Usage, etc...

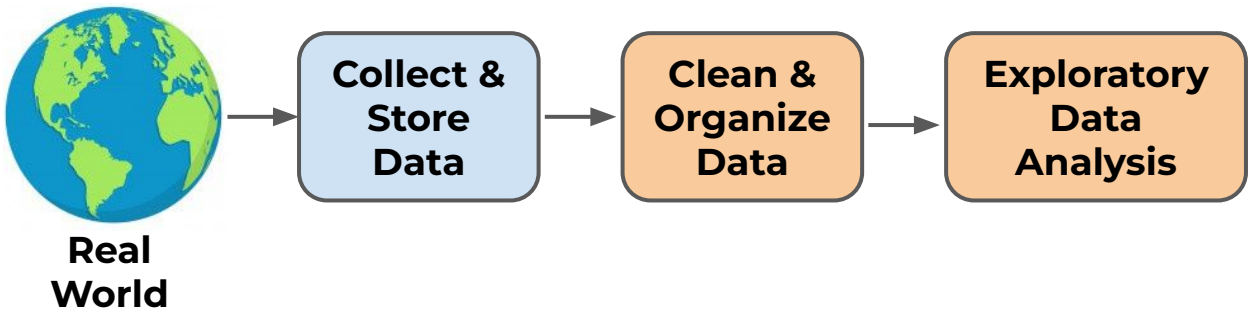
ML Pathway


Raw Data
Process & Store Data
Real World
SQL Database, CSV files, Excel, Cloud Storage

ML Pathway



ML Pathway


Real World
Collect & Store Data
Clean & Organize Data
Exploratory Data Analysis

ML Pathway

Real World

Collect & Store Data → Clean & Organize Data → Exploratory Data Analysis → Machine Learning Models

Supervised Learning:
Predict an Outcome
Unsupervised Learning:
Discover Patterns in Data

ML Pathway

Real World

Collect & Store Data

Clean & Organize Data

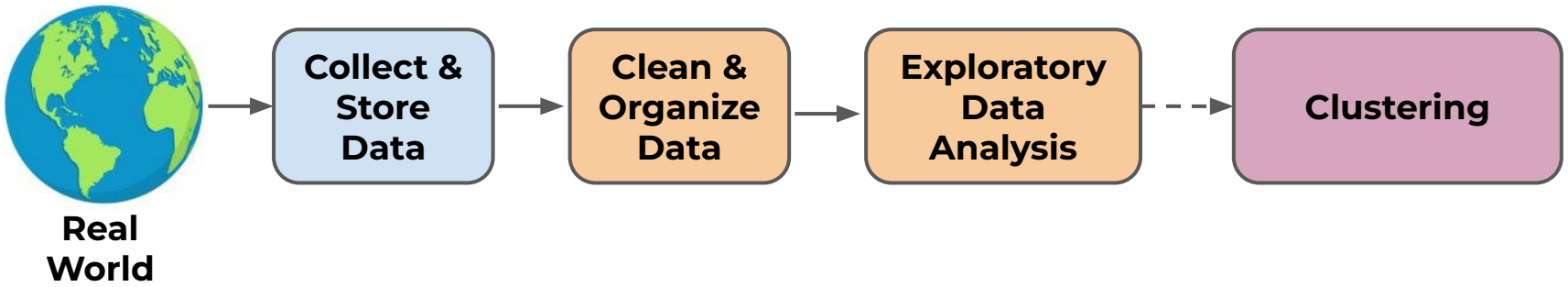
Exploratory Data Analysis

Clustering

Dimensionality Reduction

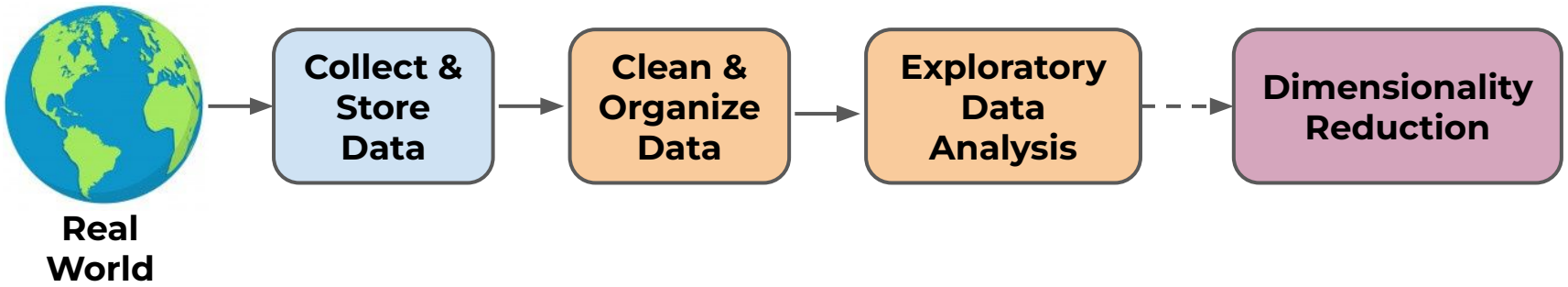
Unsupervised Learning:
Discover Patterns in Data

ML Pathway



Clustering: If we have unlabeled data, can we attempt to cluster or group similar data points together to “discover” possible labels for clusters?

ML Pathway



Dimensionality Reduction: If we have unlabeled data, can we attempt to reduce the number of features by combining them into new components? Do these new components give us further insight for the data?

Unsupervised Learning

- We'll begin by discovering clustering methods such as K-Means and Hierarchical clustering, then move on to dimensionality reduction.
- We will also learn about methods for interpreting the model results, since results and performance is much more nuanced in unsupervised learning.

Unsupervised Learning

- Questions to keep in mind:
  - *What does it really mean to “discover” labels through clustering?*
  - *Without known labels how do we measure performance?*
  - *Do combinations of features hold important insights?*

Let’s get started!