

Types of ML

Types of Machine Learning Techniques

- The machine learning algorithms which we will be covering are
 - Supervised learning algorithms
 - Unsupervised learning algorithms

Models for Supervised Learning

- When the response variable is numerical, predictive modeling is called **Regression**.
- When the response variable is nominal / categorical, predictive modeling is called **Classification**. The values of the response variable can be considered as “class labels” in this case.

Examples: Supervised Learning

- Naïve Bayes
- K-NN
- Decision Trees
- Regression Models
- Neural Nets
- Support Vector Machines

Partitioning in Supervised Learning

- In Supervised Learning, we partition the data
- We typically deal with two or three partitions:
 - a training set,
 - a validation set,
 - and sometimes an additional test set.

Training Partition

- Typically the largest partition
- Contains the data used to build the various models we are examining
- Generally used to develop multiple models.

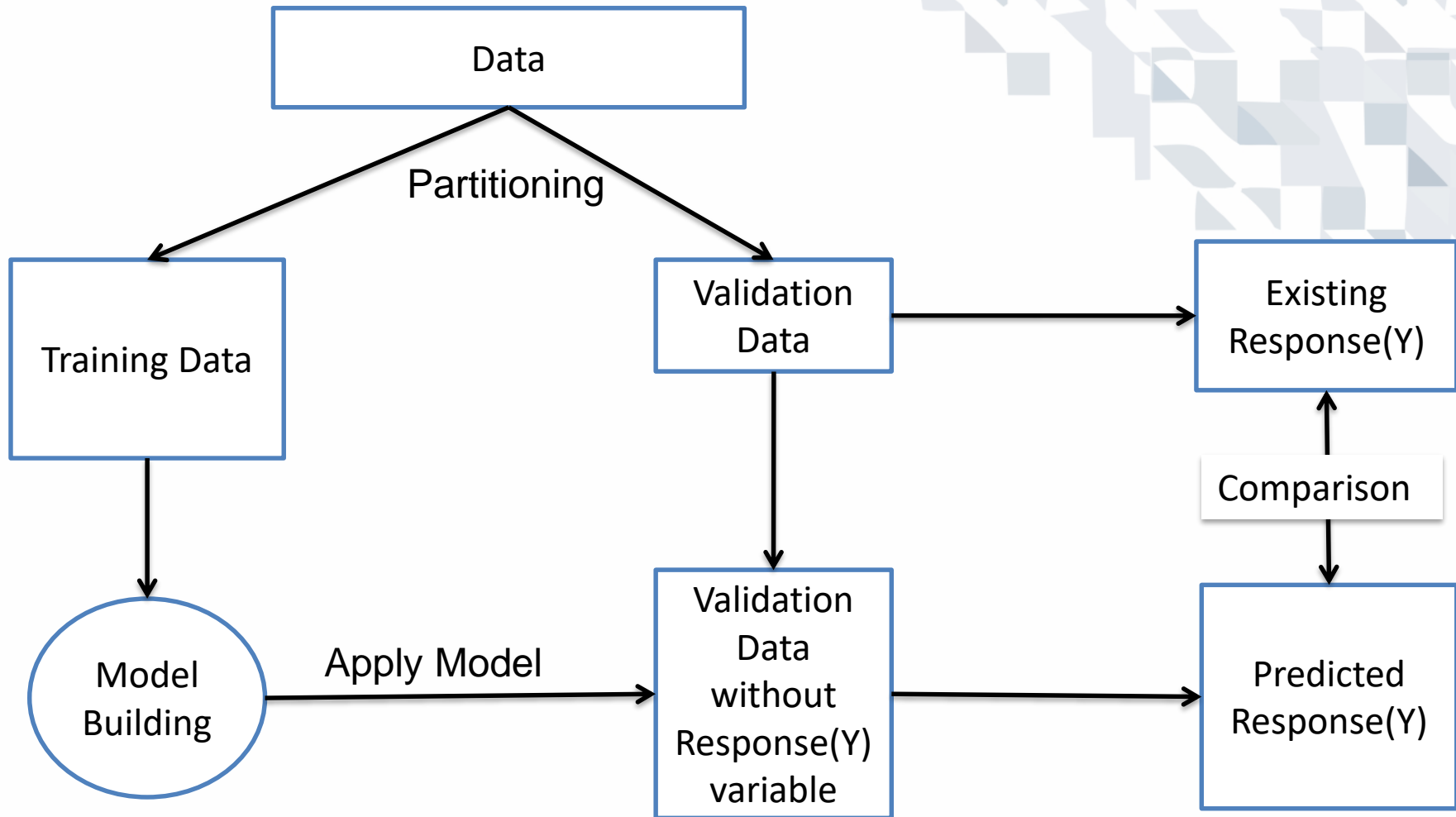
Validation Partition

- Used to assess the performance of each model so that you can compare models and pick the best one.
- This partition is used for internally verifying the performance of the models
- Important for measuring the goodness of fit

Test Partition

- Used if we need to assess the performance of the chosen model with new data
- Used to overcome the overfitting problem

Supervised Learning Process with 2 partitions



Unsupervised Learning

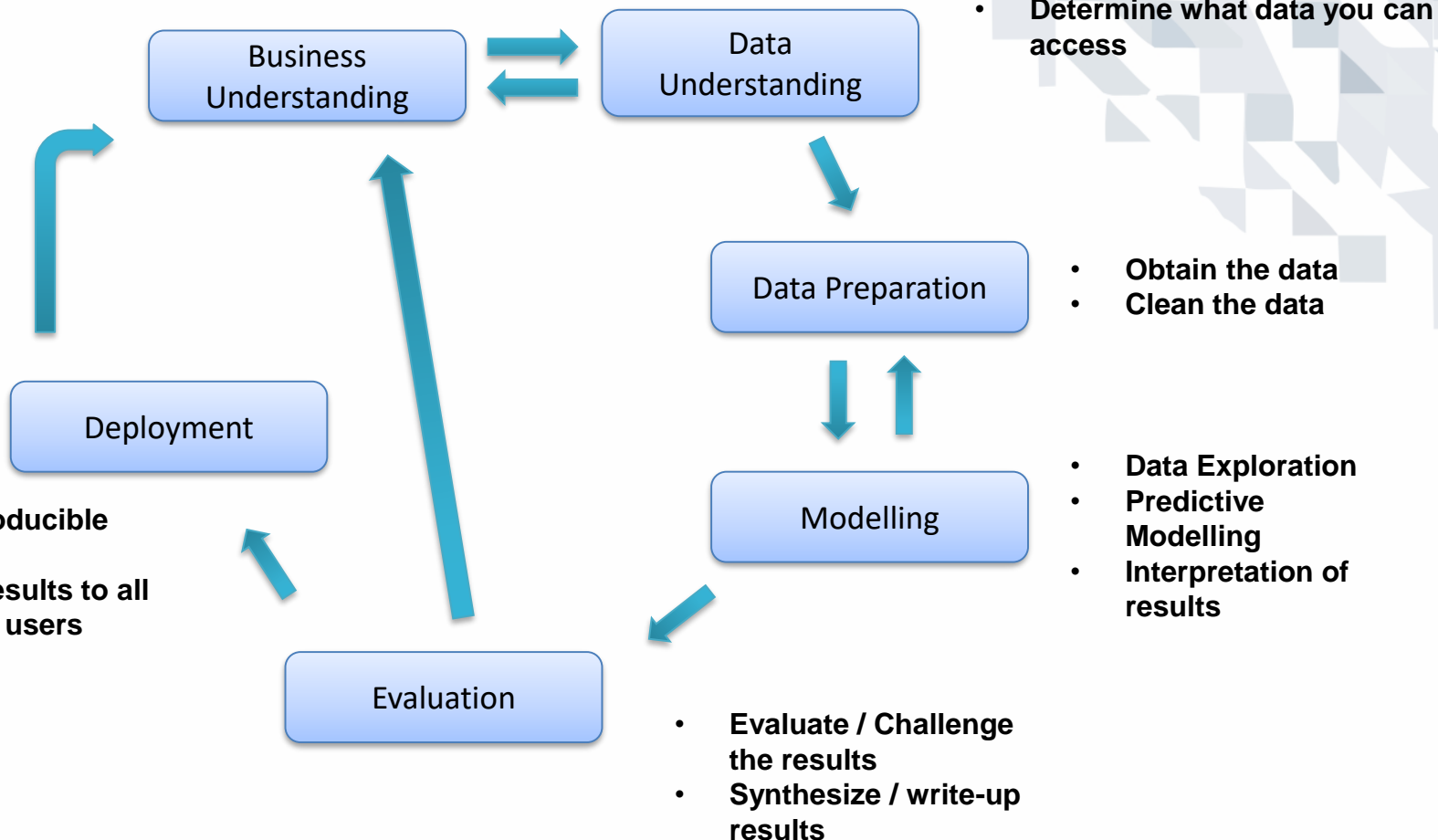
- Unsupervised learning algorithms are those used where there is no outcome variable to predict or classify.
- Association rules, data reduction methods, and clustering techniques are all unsupervised learning methods.

Examples of Unsupervised Learning

- Clustering Techniques
 - Hierarchical
 - K-means
 - K-medoids
- Principal Component Analysis
- Association Rules

Analytics Lifecycle (CRISP-DM)

- **Define a question**



Popular (Desktop-In Memory Processing) Software for Machine Learning Algorithms

- Programming
 - R
 - Python
 - Scala (Spark MLlib)
- GUI based
 - RapidMiner (Only Academic Free)
 - Weka (Free)
 - KNIME (Free)
 - SAS Enterprise Miner
 - IBM SPSS Modeller

Cloud-based Solutions for Machine Learning

- IBM Watson
- Microsoft Azure ML Studio
- Google Analytics Platform
- BigML
- Amazon Web Services