



Samsung Innovation Campus

| Artificial Intelligence Course

Mithilesh Singh
Corporate Trainer Mentor
Mob.8800686128



Chapter 1.

Introduction to Artificial Intelligence



Artificial Intelligence Course

Chapter Description



Chapter objectives

- ✓ Understand the concepts of artificial intelligence and machine learning, and learn about the types, procedures, and limitations of machine learning-based data analyses.
- ✓ Be able to analyze the internal and external environment of artificial intelligence-based corporations. And be able to analyze the commercialize business models for expanding the scope of artificial intelligence service applications in artificial intelligence-based industries.
- ✓ Be able to analyze trends in artificial intelligence-related technologies and changes in market demand.
- ✓ Learn about the entire artificial intelligence course roadmap.



Chapter contents

- ✓ **Unit 1. The Concept of Artificial Intelligence**
- ✓ **Unit 2. Applications of Artificial Intelligence**
- ✓ **Unit 3. Techniques in Artificial Intelligence**
- ✓ **Unit 4. Artificial Intelligence: Trends and Markets**
- ✓ **Unit 5. Course Roadmap**

Unit 1.

The Concept of Artificial Intelligence

- | 1.1. **Definition of Artificial Intelligence**
- | 1.2. Types and Subsets of Artificial Intelligence
- | 1.3. Definition of Machine Learning
- | 1.4. Disciplines Related to Machine Learning
- | 1.5. Types and Choices of Machine Learning-based Data Analysis
- | 1.6. Procedures for Machine Learning-based Data Analysis
- | 1.7. Reasons For Machine Learning
- | 1.8. Limitations of Machine Learning

What is Artificial Intelligence

| Definition



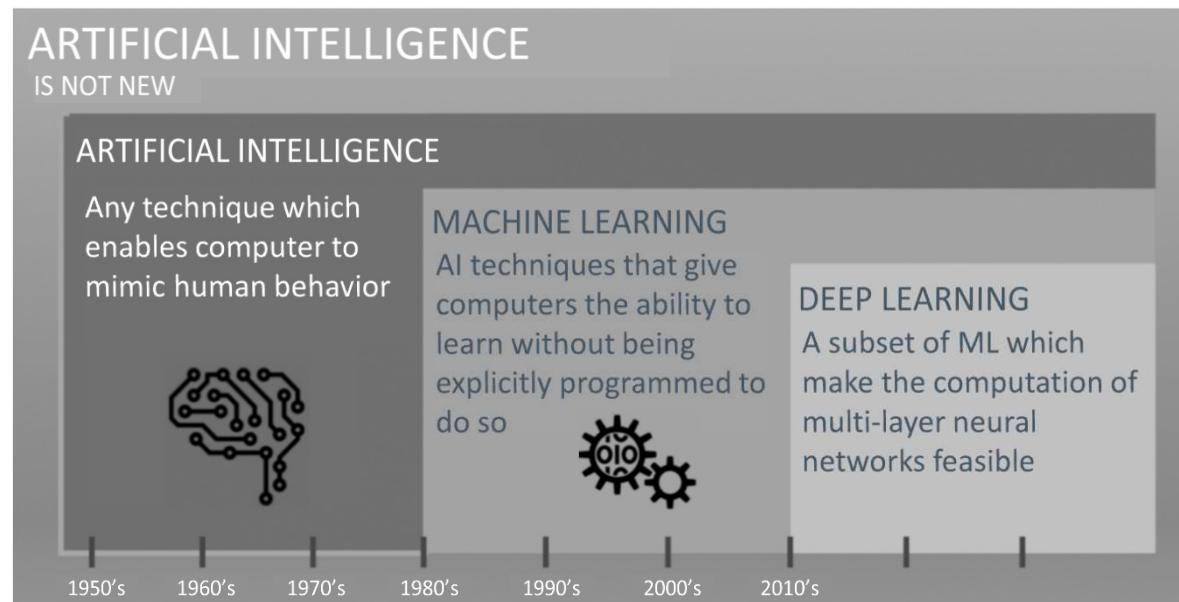
John McCarthy(1927–2011)

- ▶ **Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study that tries to make computers "smart." They work on their own without being encoded with commands. John McCarthy came up with the name "artificial intelligence" in 1955.**
- ▶ **These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions), and self-correction.**

<https://www.wired.com/2011/10/john-mccarthy-father-of-ai-and-lisp-dies-at-84/>

Building an Intelligent System that Transforms Data into Knowledge

- ▶ There is an abundance of structured and unstructured data in the modern tech era.
- ▶ Machine learning emerged in the late 20th century as a sub-field of artificial intelligence (AI) related to self-learning algorithms that extract and predict knowledge from data.
- ▶ Humans manually analyze large amounts of data to induce rules and make models.
- ▶ Machine learning can gradually improve predictive models and data-based decision-making performance by extracting knowledge more efficiently from data.

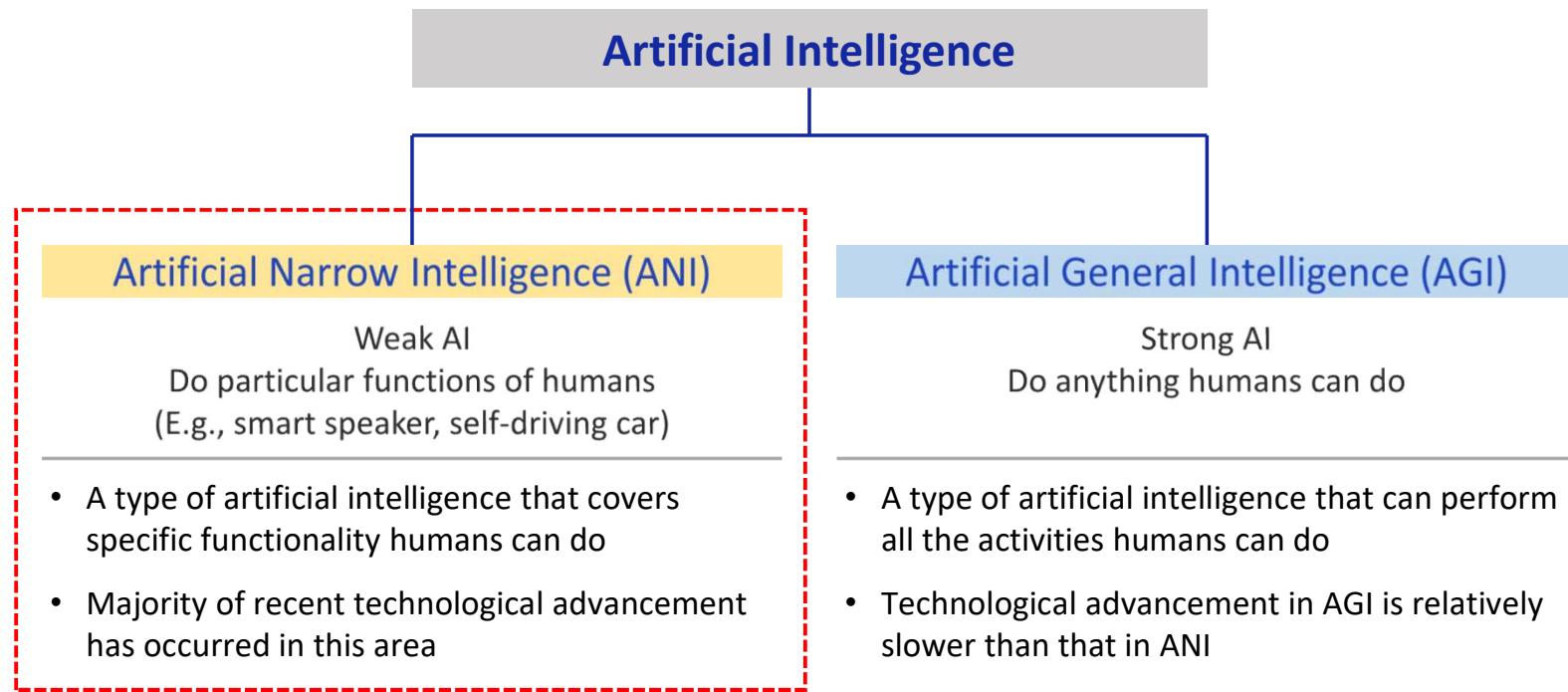


Unit 1.

The Concept of Artificial Intelligence

- | 1.1. Definition of Artificial Intelligence
- | 1.2. Types and Subsets of Artificial Intelligence
- | 1.3. Definition of Machine Learning
- | 1.4. Disciplines Related to Machine Learning
- | 1.5. Types and Choices of Machine Learning-based Data Analysis
- | 1.6. Procedures for Machine Learning-based Data Analysis
- | 1.7. Reasons For Machine Learning
- | 1.8. Limitations of Machine Learning

Types of Artificial Intelligence



Main focus of this lecture and the entire course

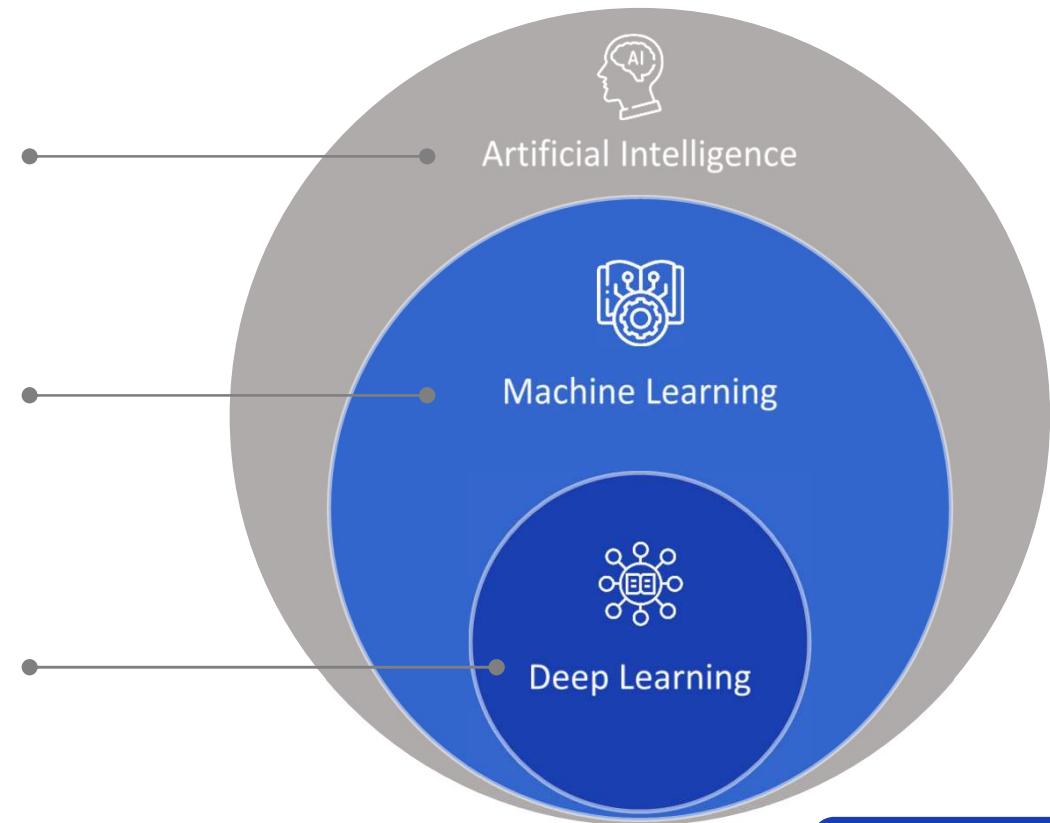
What are subsets of AI?

I Subsets of Artificial Intelligence

Artificial Intelligence
Any technique that enables computers to mimic human behavior

Machine Learning
A subset of AI techniques that use statistical methods to enable machines to improve through experiences

Deep Learning
A subset of ML that makes the computation of multi-layer neural networks feasible



Source: KD Nuggets

Unit 1.

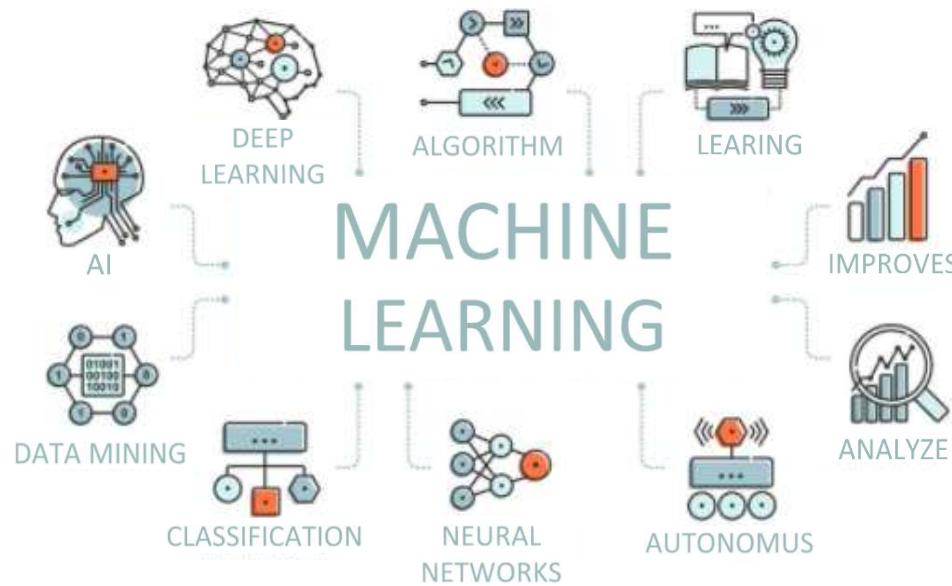
The Concept of Artificial Intelligence

- | 1.1. Definition of Artificial Intelligence
- | 1.2. Types and Subsets of Artificial Intelligence
- | **1.3. Definition of Machine Learning**
- | 1.4. Disciplines Related to Machine Learning
- | 1.5. Types and Choices of Machine Learning-based Data Analysis
- | 1.6. Procedures for Machine Learning-based Data Analysis
- | 1.7. Reasons For Machine Learning
- | 1.8. Limitations of Machine Learning

Definition of Machine Learning

| **Machine learning** is a field of **artificial intelligence** and the study of computer algorithms that automatically improve through examples and experiences.

- ▶ “A field of research that develops algorithms that allow machines to learn from data and execute actions that are not explicitly specified by code” - Arthur Samuel, 1959
- ▶ “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improved with experience E.” - Tom Mitchell, 1977



- | “A field of research that develops algorithms that allow machines to learn from data and execute actions that are not explicitly specified by code” - Arthur Samuel, 1959
- | “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improved with experience E.” - Tom Mitchell, 1977



- | The task T is to classify dogs and cats, and the performance P represents a measure of classifying dogs and cats. E can be said to be “learning” if the performance of classifying dogs and cats gradually improves through experience, or data (10,000 photos).

Unit 1.

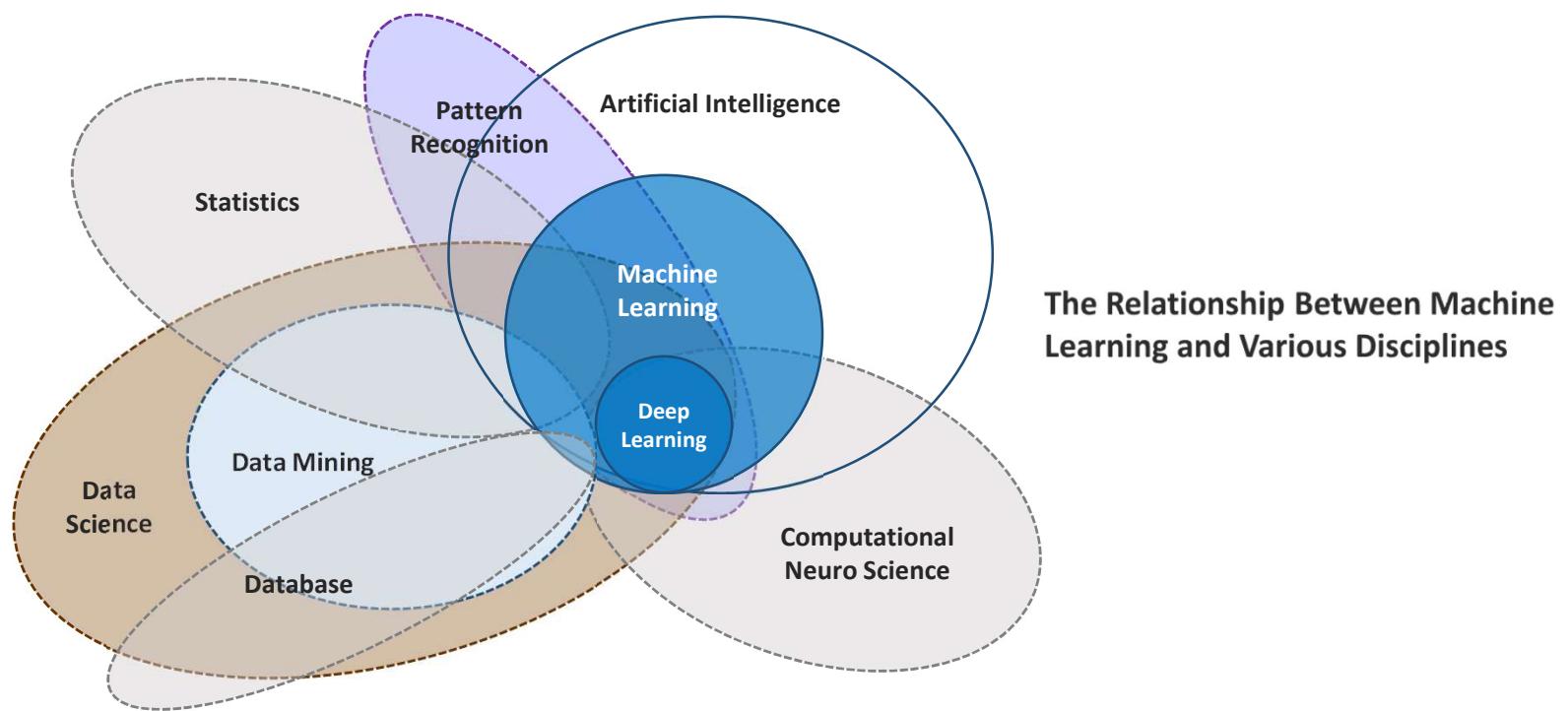
The Concept of Artificial Intelligence

- | 1.1. Definition of Artificial Intelligence
- | 1.2. Types and Subsets of Artificial Intelligence
- | 1.3. Definition of Machine Learning
- | 1.4. Disciplines Related to Machine Learning
- | 1.5. Types and Choices of Machine Learning-based Data Analysis
- | 1.6. Procedures for Machine Learning-based Data Analysis
- | 1.7. Reasons For Machine Learning
- | 1.8. Limitations of Machine Learning

Disciplines Related to Machine Learning

I Studies Related to Machine Learning

- ▶ Machine learning is an interdisciplinary study that combines academic backgrounds and achievements in various fields rather than limited to the technology or methodology in one area. These fields include probability and statistics, computer science, database theory, cognitive science, neuroscience, and pattern recognition.



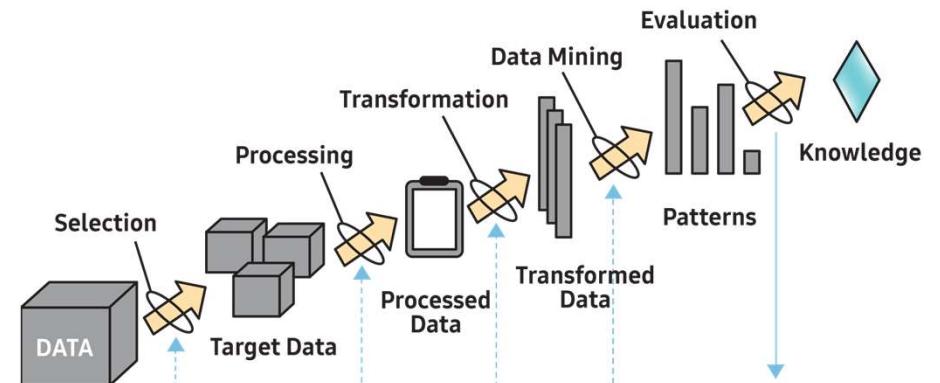
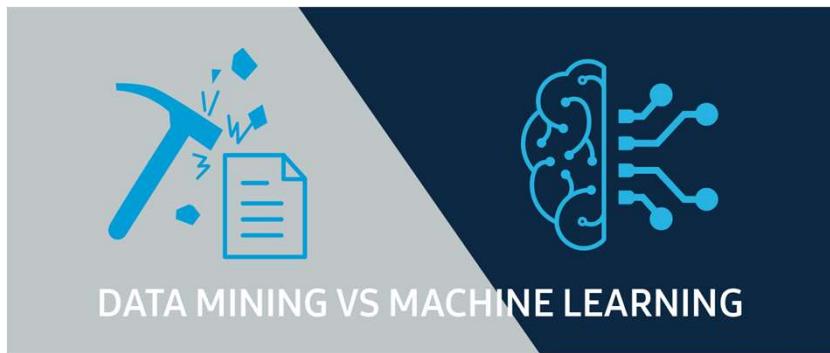
| Machine Learning vs. Statistics

- ▶ **Statistics** are traditionally regarded as the theoretical foundation that provides a scientific and systematic method for converting data into information. The areas particularly emphasized in statistics are inference and verification. Various methodologies and theories have been established to **explain how well the given data conform to the researcher's hypothesis and theory** (or how accurately the values derived from observations estimate the actual population parameters).
- ▶ Meanwhile, **machine learning** is mainly used to **solve tasks that are difficult to design or program explicit algorithms**. Most machine learning algorithms are first used to **quantify complex relationships by identifying the feature of potential mechanisms generated by data** and then to **make predictions on new data using this identified pattern**.
- ▶ At first glance, the approaches of statistics and machine learning may look the complete opposite. But except for some aspects, the methodologies that form the basis of each discipline are very similar. In fact, it can be said that **many methodologies of machine learning are based on statistical learning from statistics**.



| Machine Learning vs. Data Mining

- ▶ What is Data Mining?
 - It refers to a process of mining useful information [gold nuggets] from a large data warehouse [stone pile].
 - It is a series of processes that help companies make decisions to secure competitiveness. These processes involve finding and modeling relationships, patterns, and rules between data existing in large amounts of data.
- ▶ Machine learning is also deeply related to data mining because it extracts useful rules, knowledge expression, or judgment criteria from data.
- ▶ **Data mining** is a process of systematically and automatically discovering meaningful rules or patterns within large-scale stored data and intellectualizing them. **Machine learning** is somehow different as it is a process in which **computer programs learn, make predictions, and research and build algorithms for said process**.



| Statistical Analysis vs. Data Mining

▶ Traditional Statistical Analysis

- With a target group, analysis is conducted based on different assumptions on the distribution or model of a population.
→ The process of inferring the entire parameter of a population by observing a sample.

▶ Data Mining

- There are no prerequisites for the distribution or model that is inevitably involved in the sample survey or experiment.
→ The process of extracting necessary information or knowledge using the entire data of a population.
- Data mining requires a large amount of data.



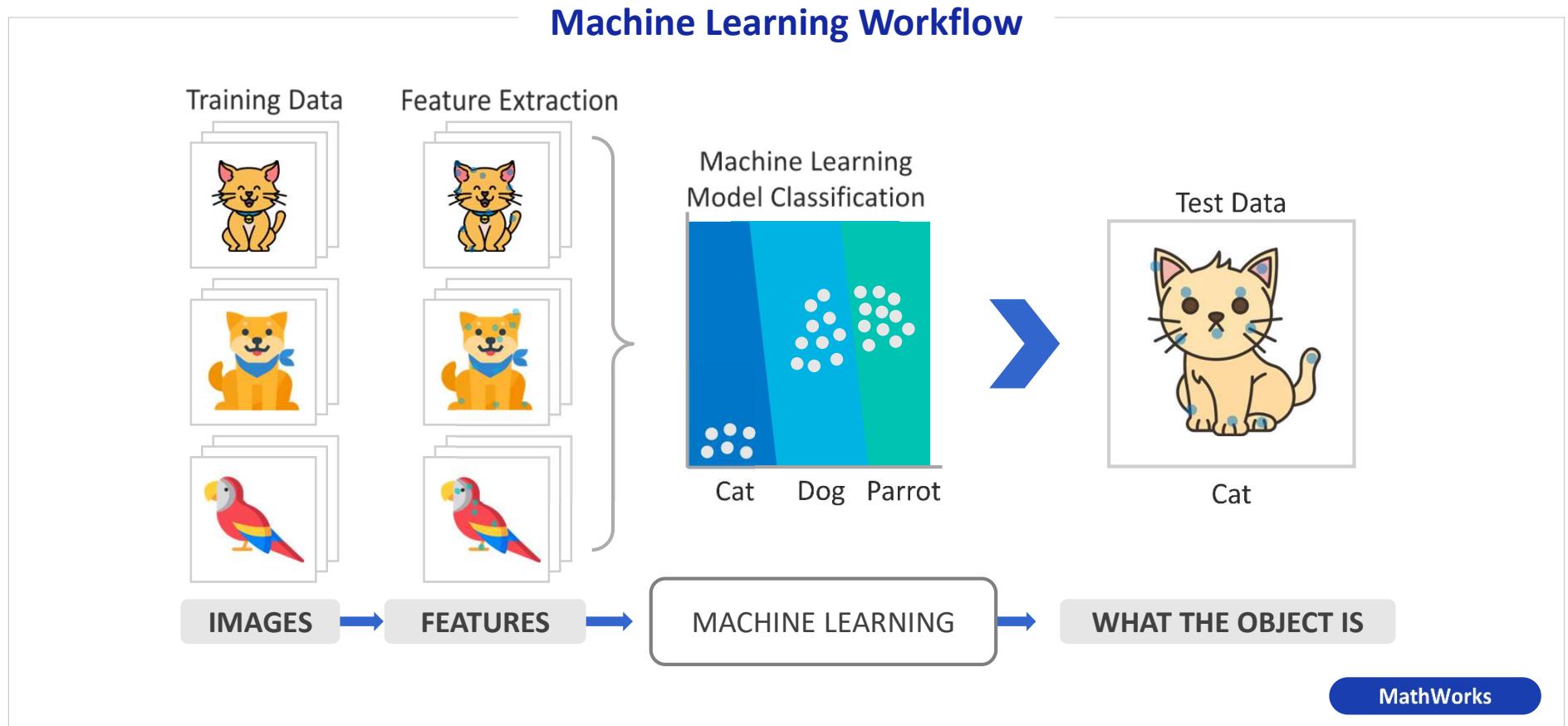
Unit 1.

The Concept of Artificial Intelligence

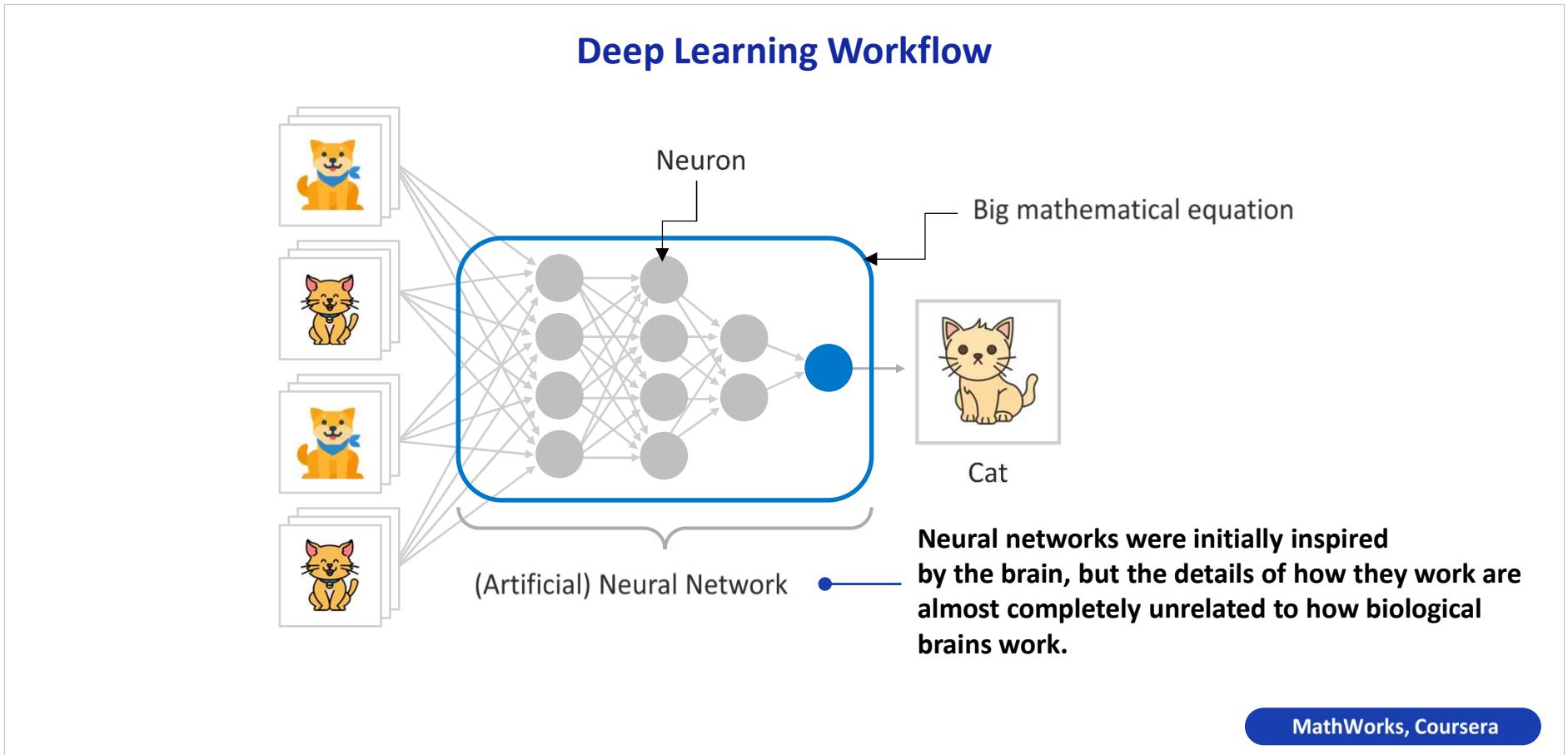
- | 1.1. Definition of Artificial Intelligence
- | 1.2. Types and Subsets of Artificial Intelligence
- | 1.3. Definition of Machine Learning
- | 1.4. Disciplines Related to Machine Learning
- | 1.5. Types and Choices of Machine Learning-based Data Analysis
- | 1.6. Procedures for Machine Learning-based Data Analysis
- | 1.7. Reasons For Machine Learning
- | 1.8. Limitations of Machine Learning

Types and Choices of Machine Learning-based Data Analysis

| Machine Learning vs. Deep Learning

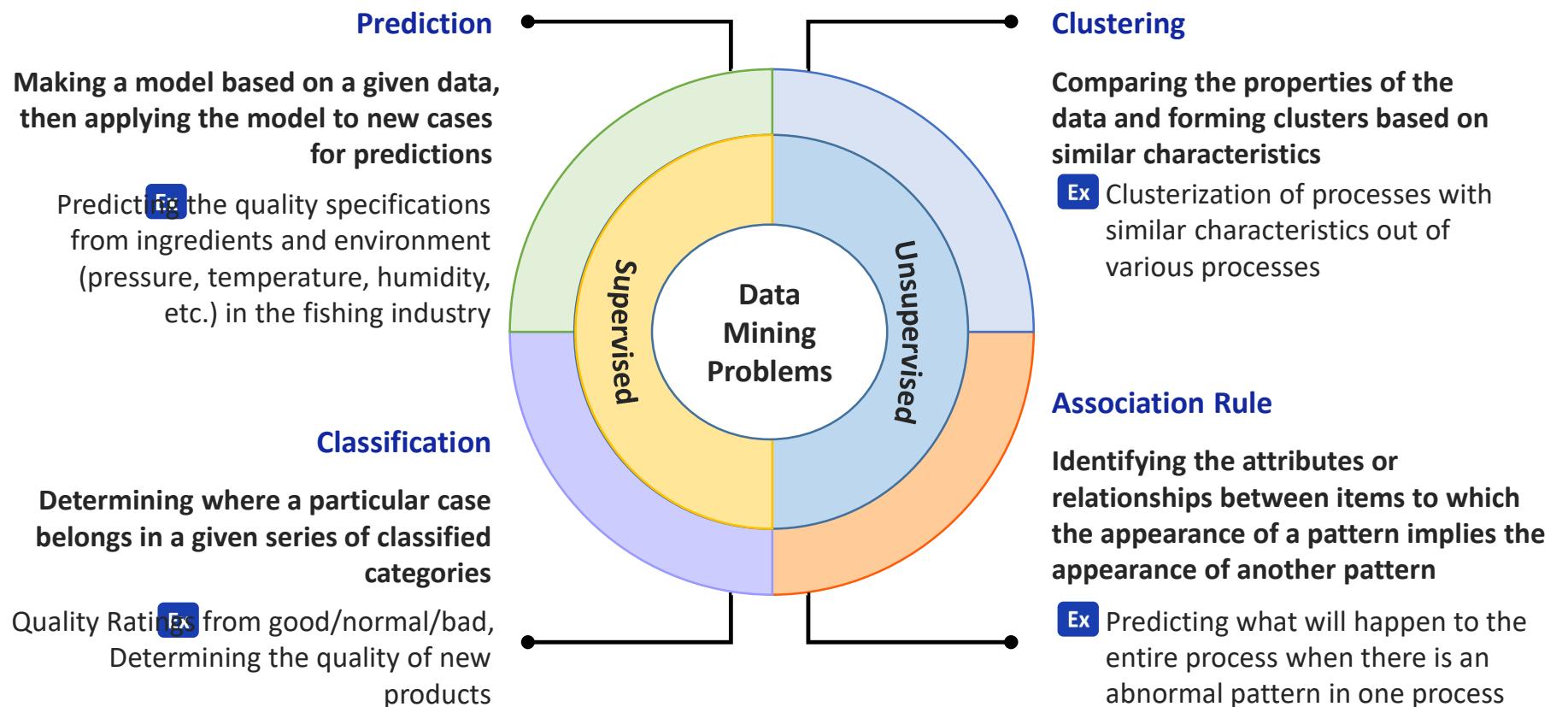


| Machine Learning vs. Deep Learning



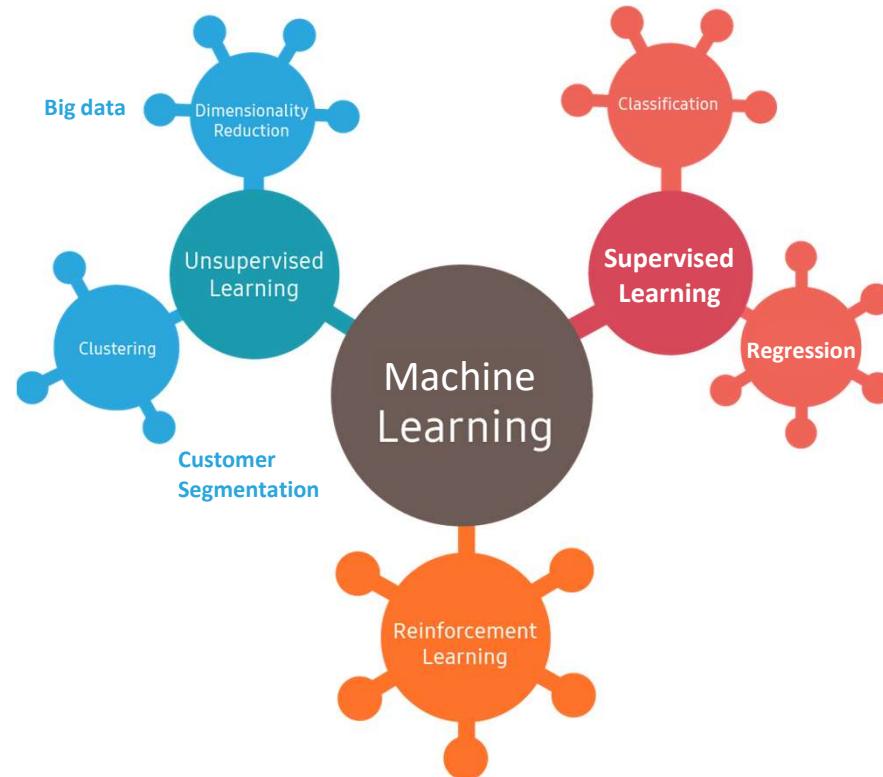
I Types of Machine Learning-based Data Analysis

- When analyzing a large amount of data with machine learning technology, patterns that were invisible outwardly can be found, called data mining. Data mining deals with the following four problems.



I Types of Machine Learning-based Data Analysis

- ▶ There are various criteria and perspectives for classifying machine learning-based data analysis techniques. But in general, they are classified into supervised learning and self-learning (or unsupervised learning), depending on the existence of objective variables (or response variables, output target values, etc.). They can also be classified as reinforcement learning and semi-supervised learning.



I Supervised Learning

- ▶ Supervised learning focuses on expressing the relationship between explanatory variables (also known as independent variables, features, etc.) and objective variables (also known as response variables, dependent variables, target variables, output values, etc.) as well as predicting future observations. It is suitable for solving problems for recognition, classification, diagnosis, and prediction.
- ▶ Main techniques in supervised learning can be reclassified into classification and numerical prediction techniques. This depends on the form of the objective variable (or response variable, dependent variable), whether it is a numerical (quantitative variable) or categorical (qualitative variable).

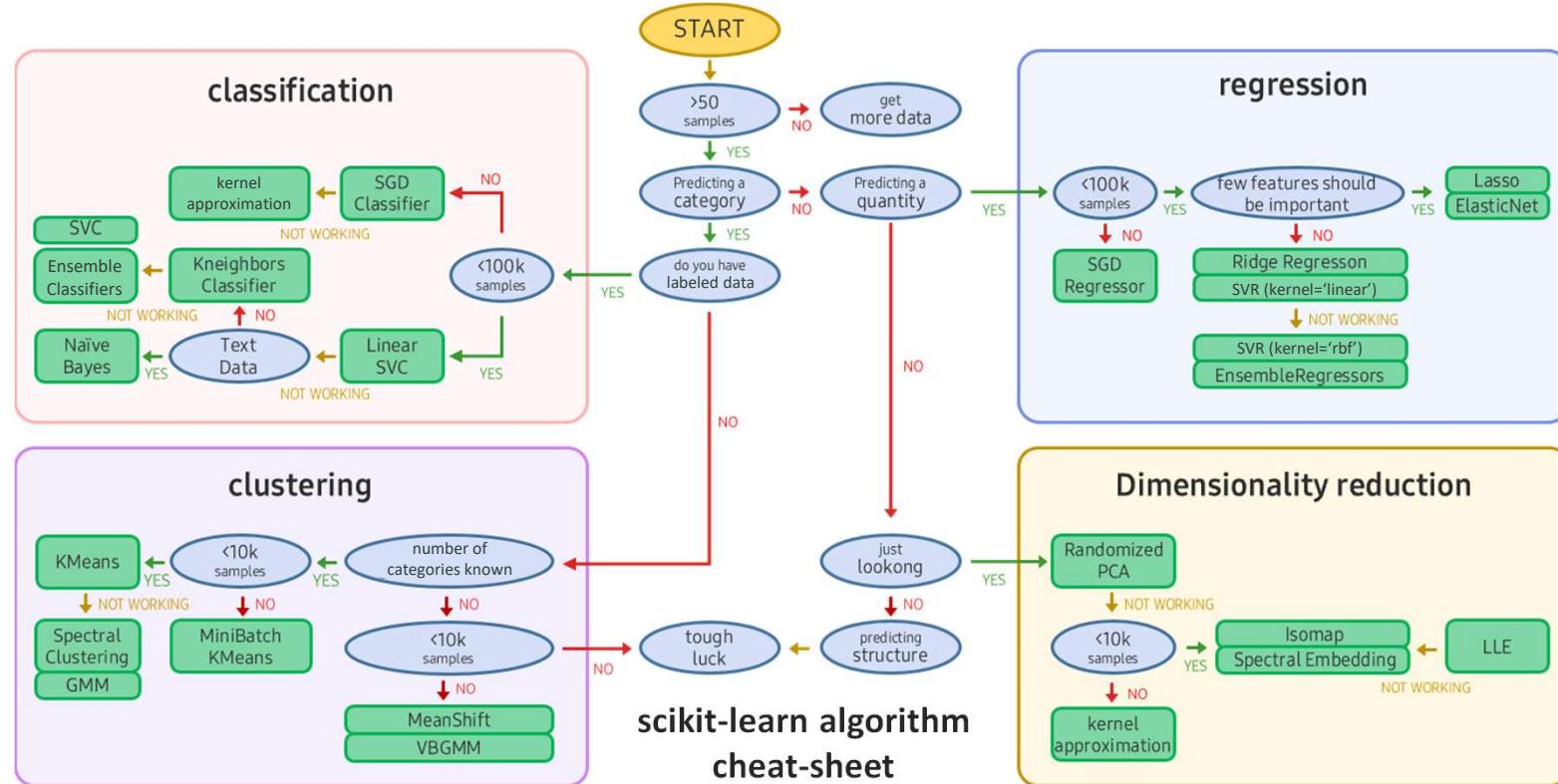
Labeled Training Set for Spam Classification (Example of Supervised Learning)



- ▶ The training data injected into the algorithm includes a desired answer called the label.

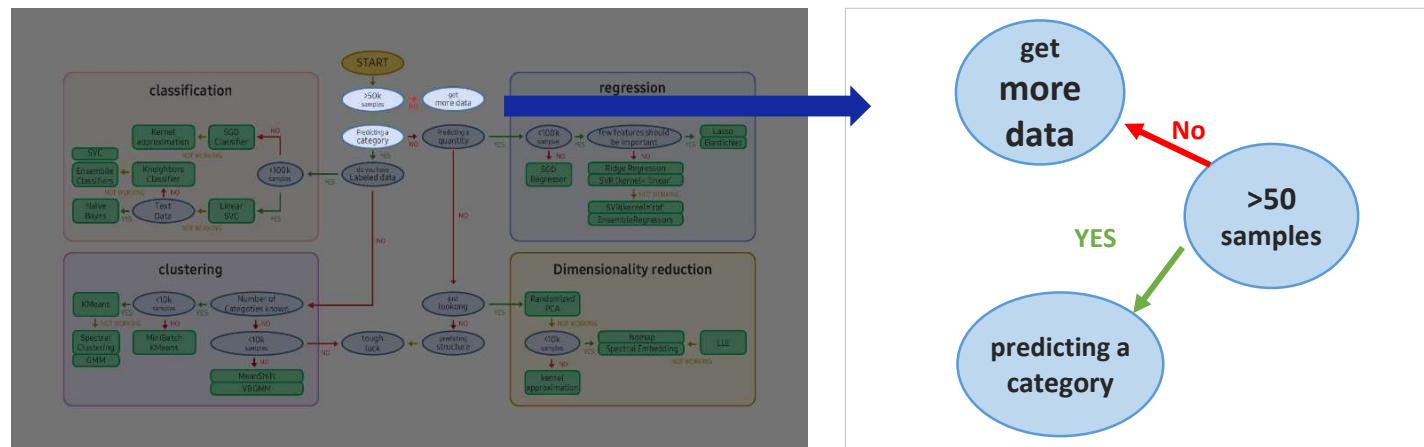
Supervised Learning

- ▶ Data mining is a combination of statistics and machine learning (or artificial intelligence). The problem to be solved by machine learning is in line with the problem of data mining.
- ▶ The figure below shows a guide to solving problems with Scikit-learn, a Python module for machine learning.



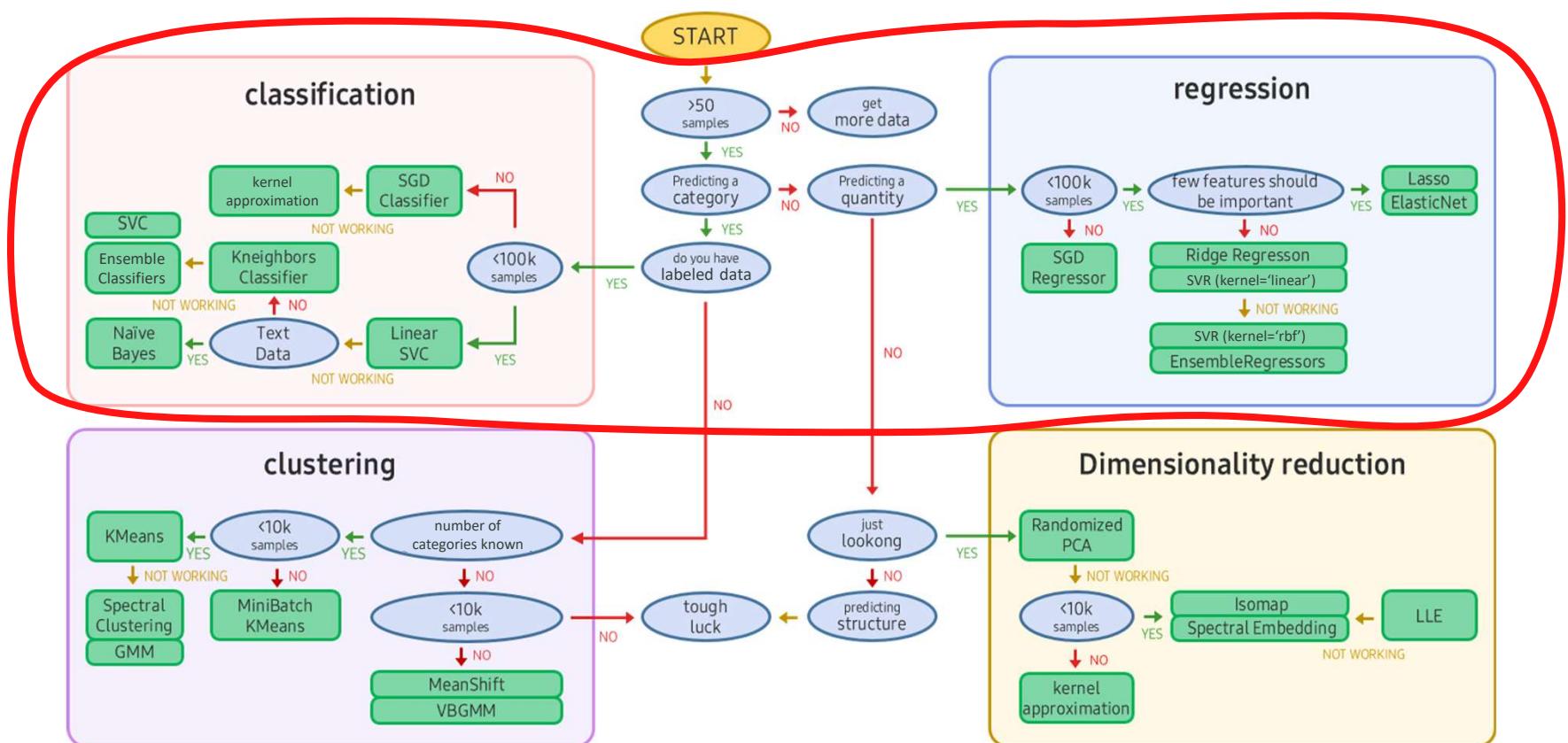
| In the Case of Insufficient Data

- ▶ Looking at the figure below, statistics-based analysis is inevitable when there is a small amount of data.
- ▶ In statistics, at least 30 samples are needed to estimate the population's features (parameters).
- ▶ As the number of samples increases, they become closer to the parameters.
- ▶ Some assumptions can be made based on approximately 30 data. The assumption is that if there are more than 30 samples, the data will be normally distributed based on the law of large numbers.
- ▶ In data mining or machine learning, a large amount of data (100,000 or more) is essential.



| Supervised Learning

- ▶ See the below figure. If there is an answer among the predicted questions, and the answer is a numerical type, it is a **regression**. If the answer is a categorical type, it is a **classification**.

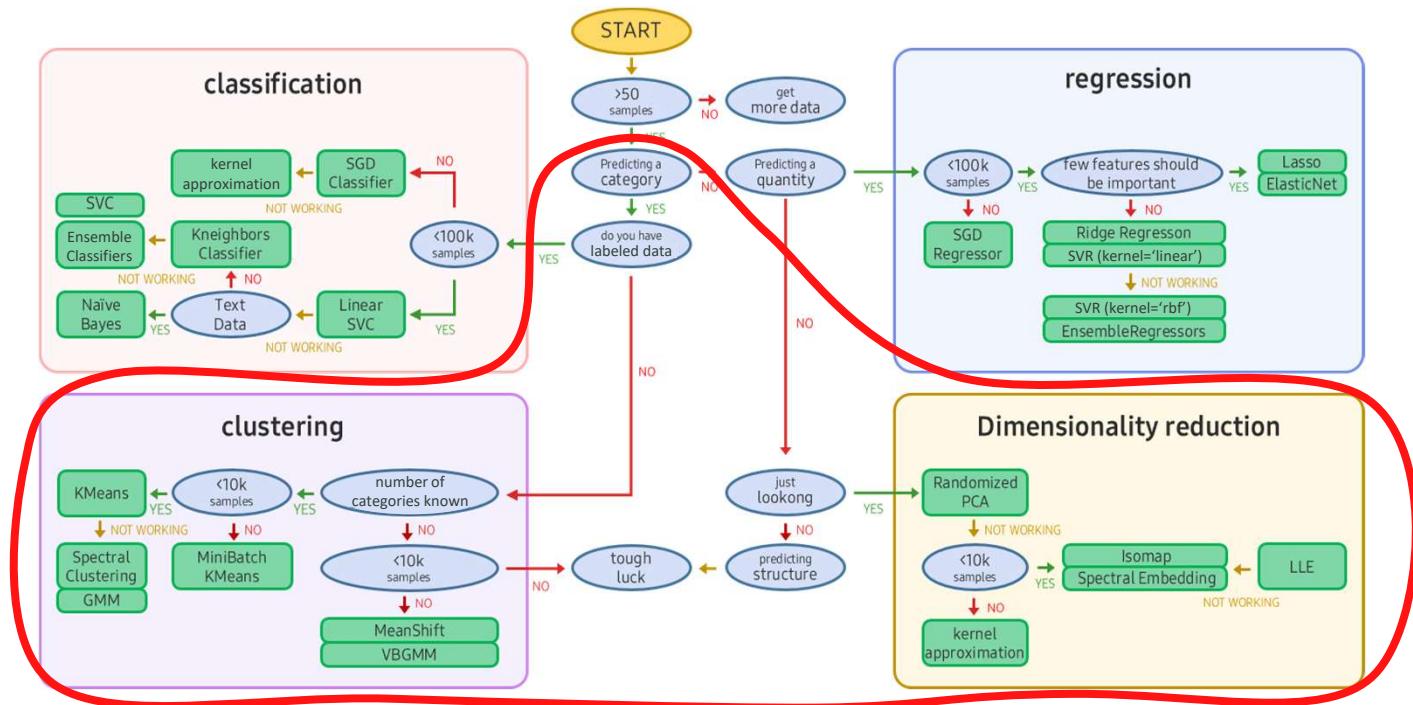


I Main Techniques and Algorithms for Supervised Learning

Type	Regression (Numerical Prediction)
K-Nearest Neighbors	Linear Regression
Logistic Regression	Extended Regression Analysis (ex: Polynomial Regression, Nonlinear Regression, Penalized Regression, etc.)
Artificial Neural Network	Artificial Neural Network
Decision Tree	Decision Tree
Support Vector Machine	Support Vector Machine (Regression)
Naïve Bayes	PLS (Partial Least Squares)
Ensemble Method (Random Forest, etc.)	Ensemble Method (Random Forest, etc.)

I Self-Learning (or Unsupervised Learning)

- ▶ **Self-learning or unsupervised learning** is a form in which learning is performed without information on objective variables (or response variables, dependent variables, target variables, and output values). It is mainly used for problems such as description, characteristic derivation, and pattern derivation.
- ▶ In general, supervised learning techniques have clear and distinct predictive purposes. Compared to those, self-learning techniques have a stronger nature of data mining to search for useful information or patterns without prior information.



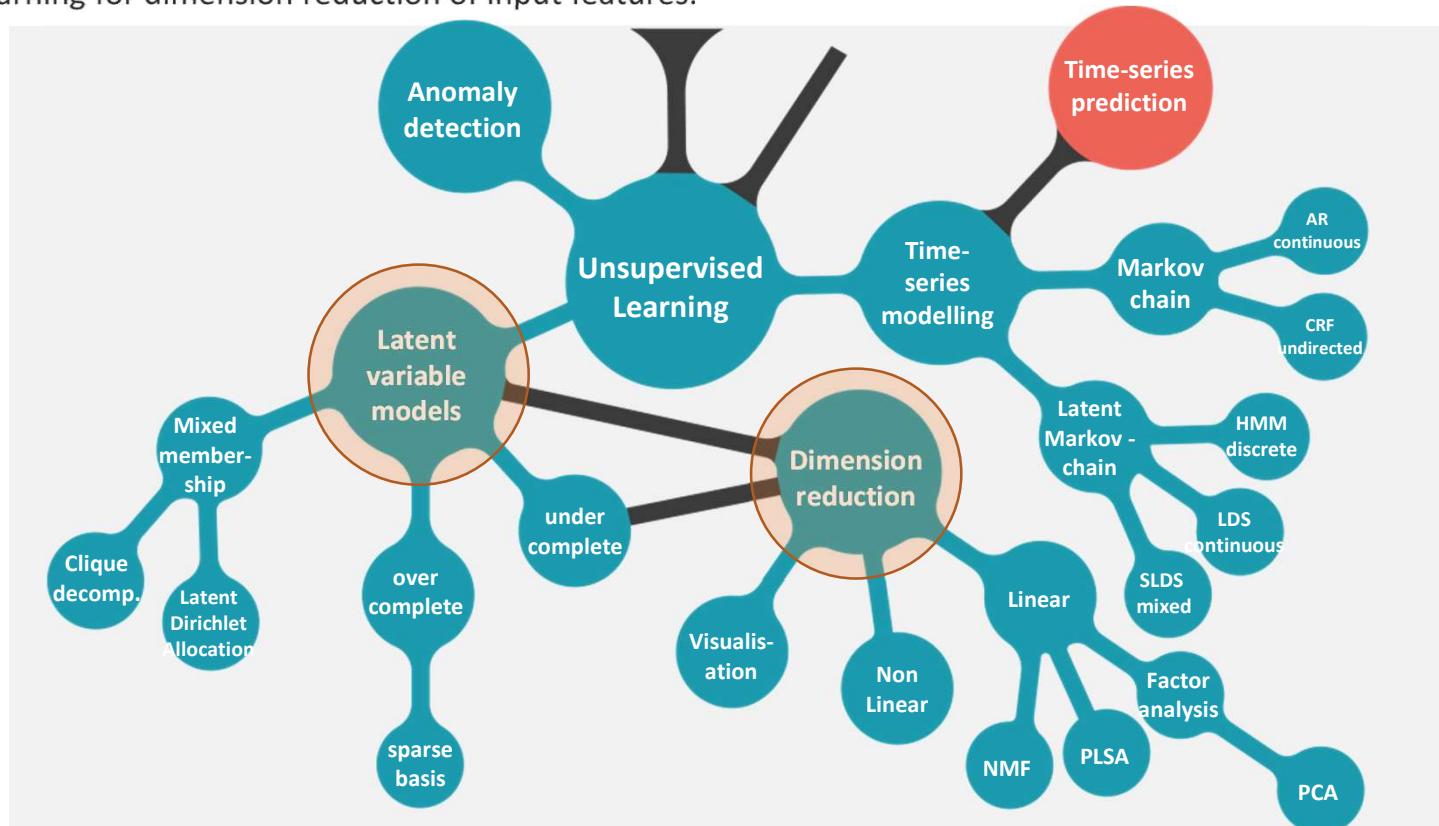
I Unsupervised Learning

- ▶ Unsupervised Learning: No label in training data, so the system must learn without any help.
- ▶ Important Unsupervised Learning Algorithms

Cluster	Visualization and Dimension Reduction	Association Rule Learning
<ul style="list-style-type: none">• K-Means• DBSCAN• Hierarchical Clustering Analysis (HCA)• Anomaly Detection and Outlier Detection• One-Class SVM• Isolation Forest	<ul style="list-style-type: none">• Principal Component Analysis(PCA)• Kernel PCA• Local Linear Embedding• t-SNE	<ul style="list-style-type: none">• Apriori• Eclat

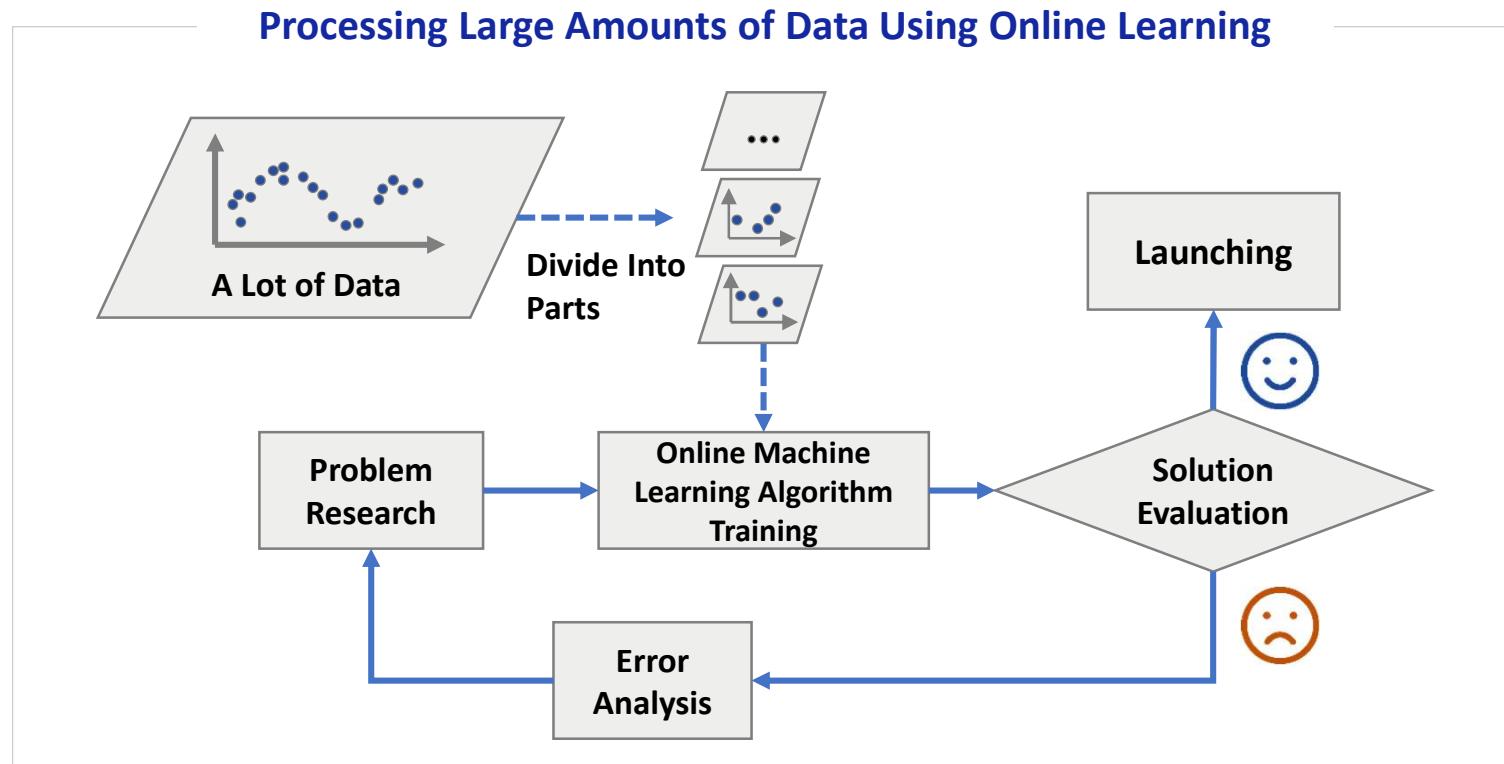
I Self-Learning (or Unsupervised Learning)

- Main techniques of self-learning, or unsupervised learning, include clustering, dimensional reduction, correlation analysis, and self-learning artificial neural network (ex. SOM). They can be used for different purposes, for example, applying them in deep learning for dimension reduction of input features.



I Batch Learning and Online Learning

- ▶ **Batch Learning:** The system cannot learn gradually.
- ▶ **Online Learning:** The system is trained by sequentially injecting data one by one or in a small batch unit called a mini-batch.



I Instance-based Learning Vs. Model-based Learning

- ▶ **Instance-based Learning:** The system learns by remembering training samples. It is generalized by comparing new data and learned samples using similarity measurements.
- ▶ **Model-based Learning:** The system creates a model from a sample and uses it for prediction.

