

Data Science - KabAI Series

A. Simon Alex

BSc(CS), MSc(CS)-MUK
simonalexatwiine@gmail.com

Mathematics & Statistics

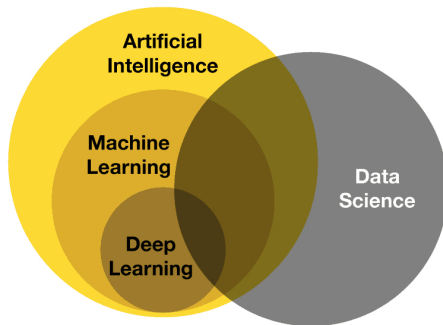


Overview

- 1 Data Science - KabAI Series
 - What is Data Science?
 - ML ToolBox
 - Mathematics & Statistics



What is Data Science?¹



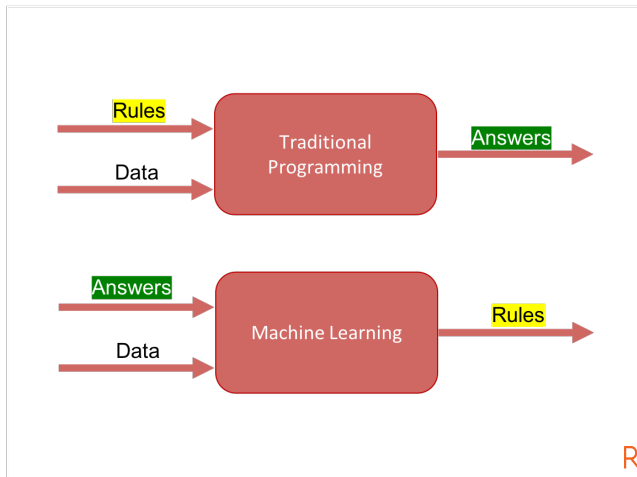
¹Data scientists **measure** whereas statisticians **estimate**

What is Artificial Intelligence?

- There have been five generations of computer evolution.
 - First generation: vacuum tubes
 - Second generation: transistors
 - Third generation: integrated circuits (ICs)
 - Fourth generation: microprocessors
 - Fifth generation: artificial intelligence



What is Machine Learning?



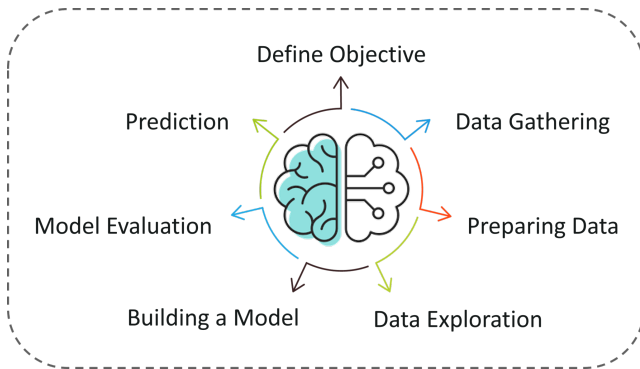
Applications of Machine Learning

- 
- Self-driving cars
 - Recommendation engines
 - Spam filtering
 - Face recognition

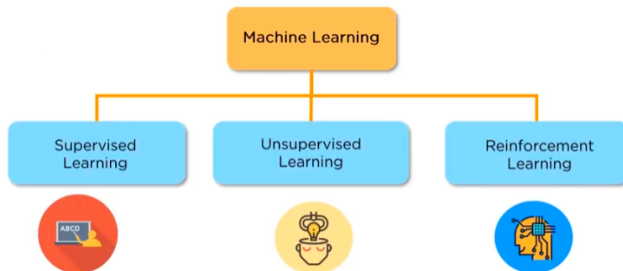
- Algorithm
- Model
- Training data
- Testing data



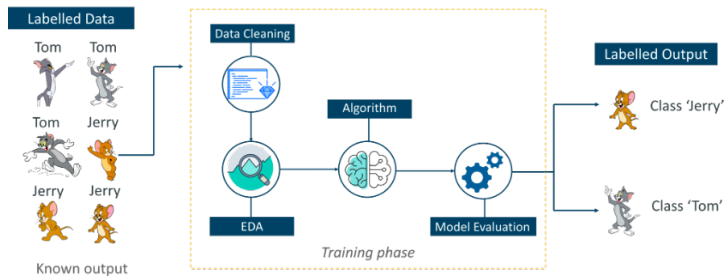
Machine Learning Process



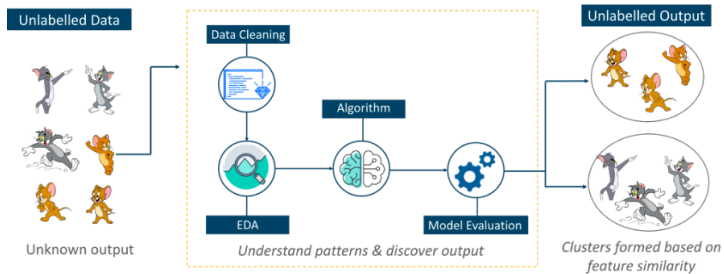
Machine Learning Types



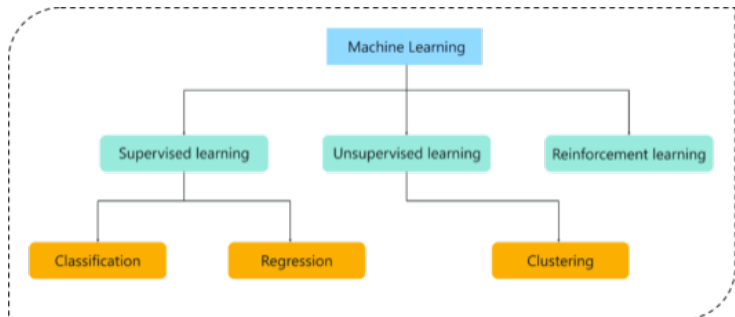
Supervised Learning



Unsupervised Learning



Types of Problems in Machine Learning



Machine Learning Vs. Deep Learning

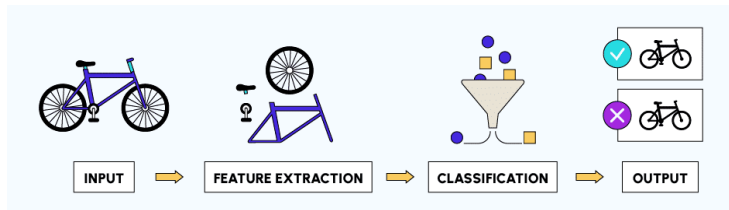


Figure: Machine Learning

Machine Learning Vs. Deep Learning

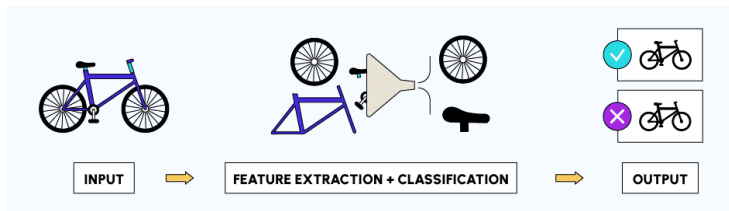
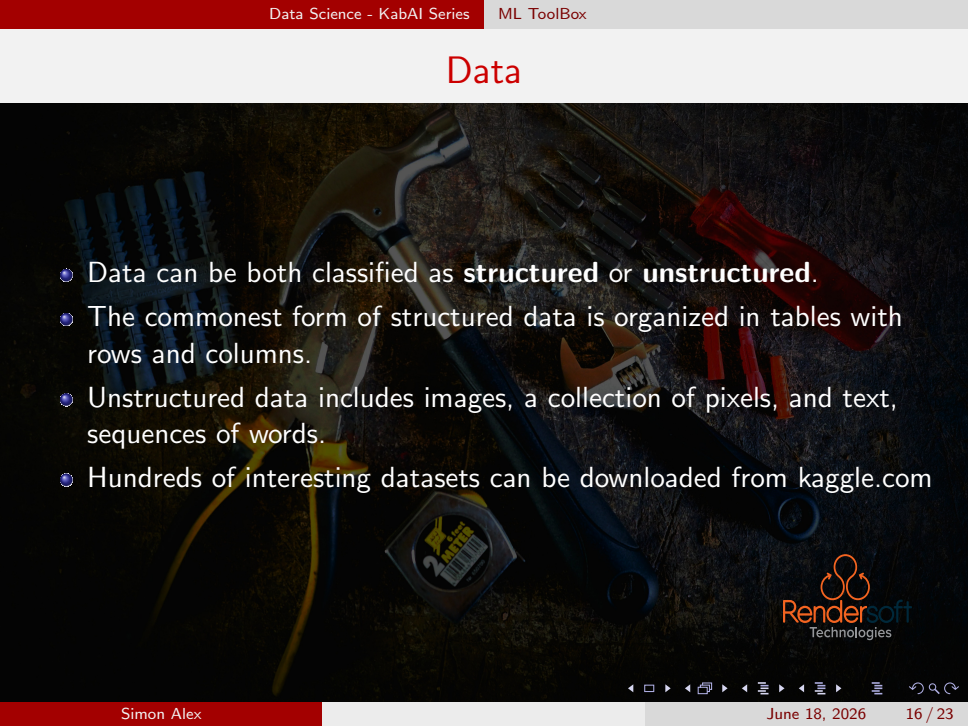


Figure: Deep Learning

ML ToolBox

- 
- Data (gold mine)
 - Infrastructure (platform & tools)
 - Algorithms (mathematics & data statistics)

Data

- 
- Data can be both classified as **structured** or **unstructured**.
 - The commonest form of structured data is organized in tables with rows and columns.
 - Unstructured data includes images, a collection of pixels, and text, sequences of words.
 - Hundreds of interesting datasets can be downloaded from [kaggle.com](https://www.kaggle.com)

Tabular Dataset's Structure

Vectors & Matrices

	Vector	Matrices	
	Feature 1	Feature 2	Feature 3
Row 1			
Row 2			
Row 3			
Row 4			

Tabular Dataset's Structure

Predictor & Response

	Vector	Matrices		
	Maker (X)	Year (X)	Model (X)	Price (y)
Row 1				
Row 2				
Row 3				
Row 4				

Infrastructure

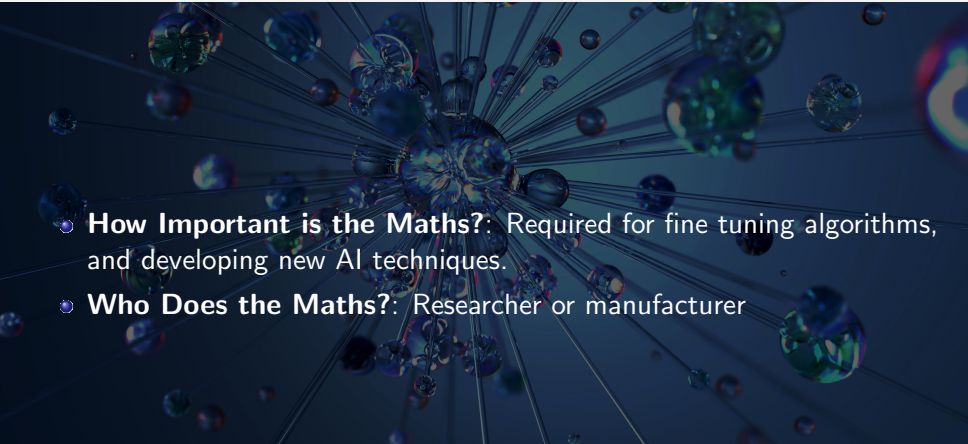
Languages	Platform	Libraries
PYTHON C, C++ R MATLAB Octave	Anaconda, Google Colab	Numpy, Pandas, Matplotlib



Algorithms

- Linear regression
- Logistic regression
- Decision trees
- K-nearest neighbors

Mathematics & Statistics

- 
- **How Important is the Maths?:** Required for fine tuning algorithms, and developing new AI techniques.
 - **Who Does the Maths?:** Researcher or manufacturer

References

- Dhinakaran, M., et al. Machine Learning For Absolute Beginners. Booksclinic Publishing, 2023.
- Papoulis, Athanasios. Probability and statistics. Prentice-Hall, Inc., 1990.
- Bruce, Peter, Andrew Bruce, and Peter Gedeck. Practical statistics for data scientists. O'Reilly Media, 2020.



Questions?

