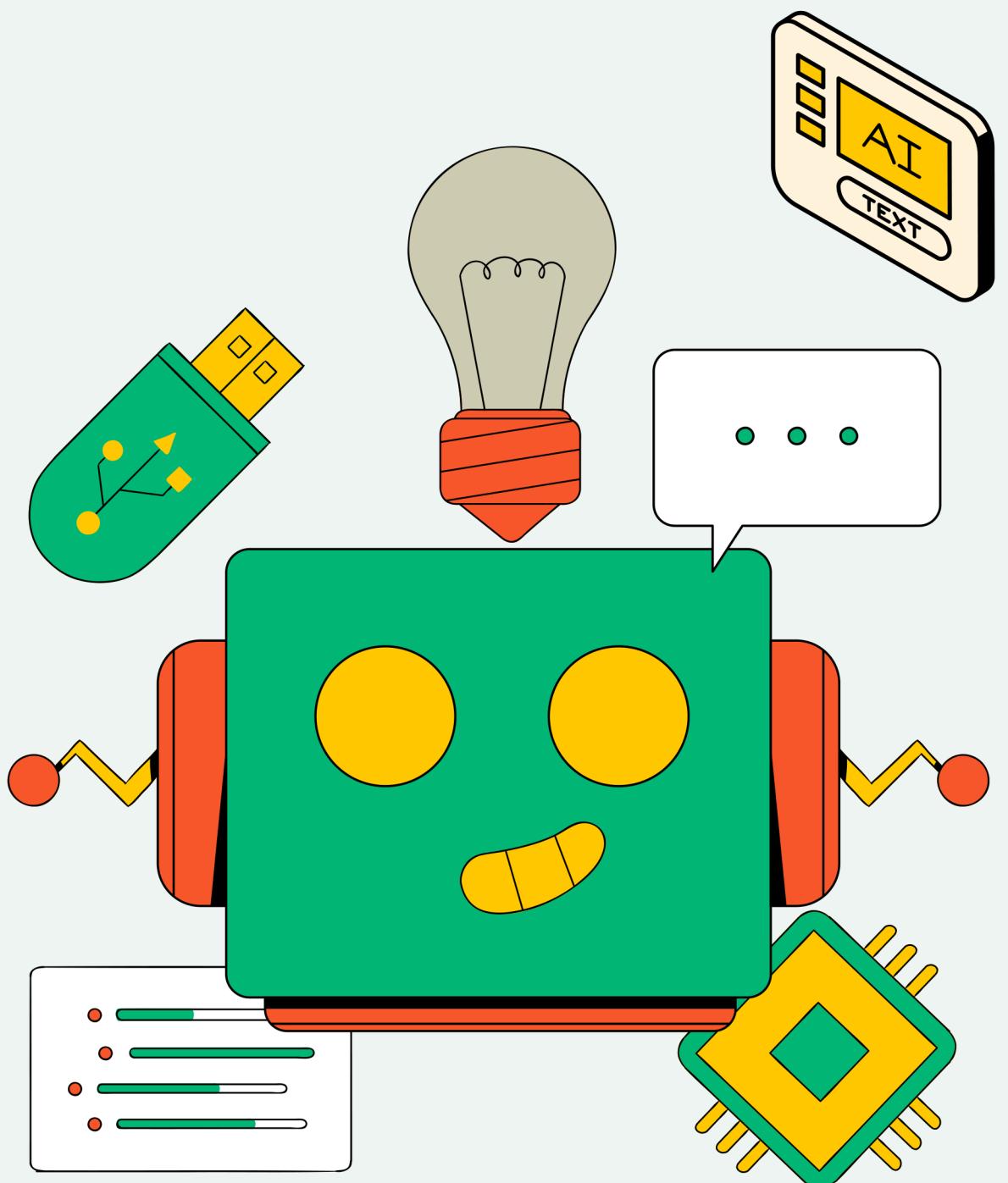




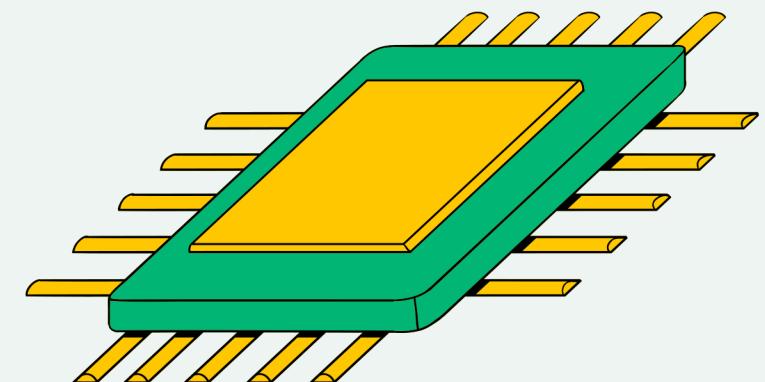
DCSE SUMMER COURSE ON
AI AND ROBOTICS



SIMPLE MACHINE LEARNING ON UNSTRUCTURED DATA PRESENTATION

PRESENTED BY:

LOUIS WIDI



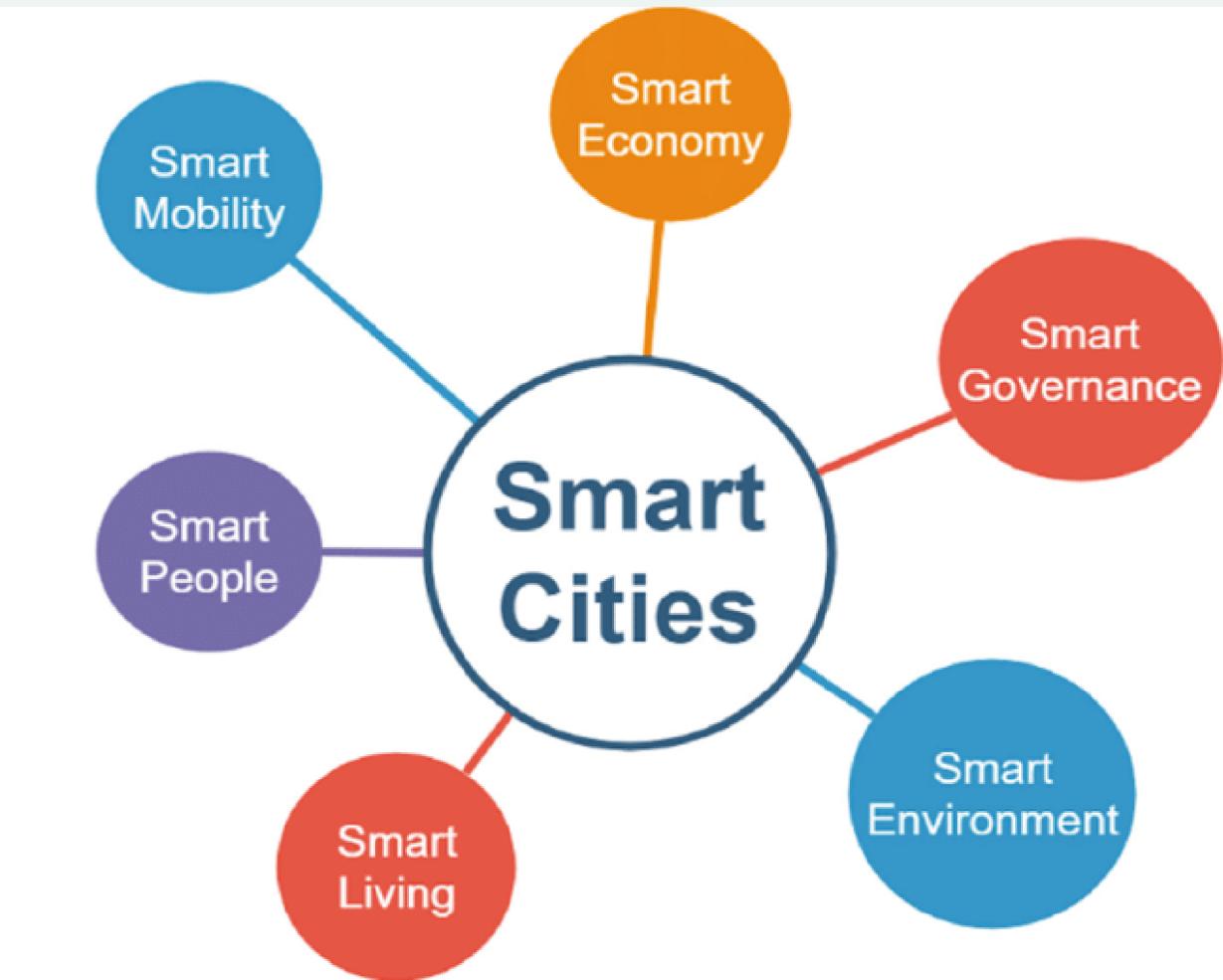
TUTORIAL OUTLINE

- Revisit Smart City
- Today's goal, spam reporting
- The idea of using unstructured data in AI
- Representing the information
- Approach on utilizing the information
- Approximation models
- Hands on practice
- Evaluations and Discussion
- Real World Applications



REVISIT SMART CITY

“A concept to use **information, communication, and technology** to provide services and solve city problems”



6 Pillars of Smart City (source: [Matos et al., 2019](#)).

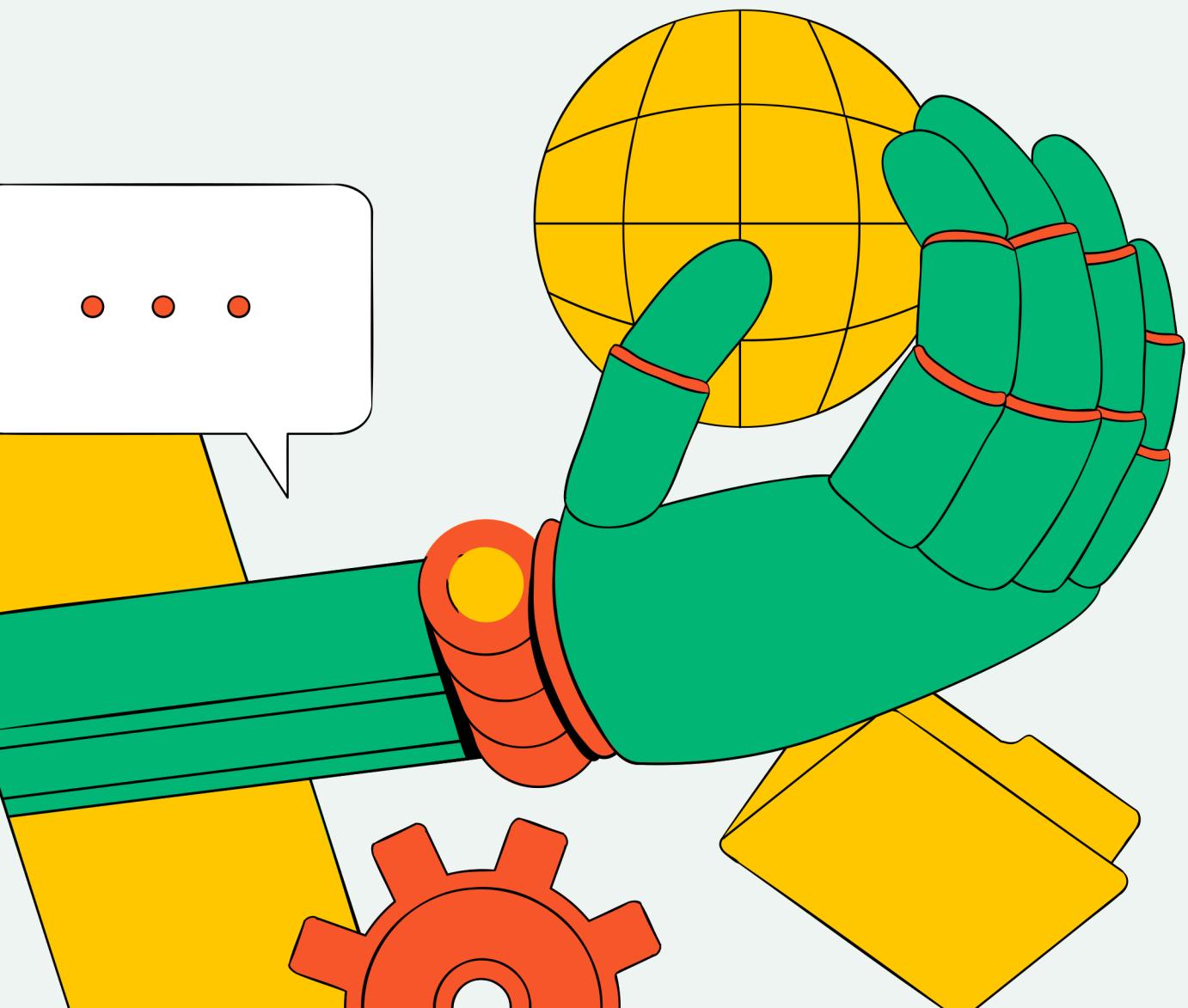
There are people who **live in the city** and there are people who **govern the city**. Smart city is a way of thinking that integrates technologies for both parties living their life.

Implementations can include:

- Traffic Monitoring
- Illegal Parking Detection
- Spam Message Reporting
- Healthcare Management System
- Face Recognition in Public Facilities



TODAY'S TUTORIAL



We will try to implement a simple spam detection mechanism using machine learning which includes:

- 1.Understanding the data
- 2.Extracting information from the data on a mathematical manner
- 3.Creating a machine learning model to
- 4.Evaluating the model



PRELIMINARIES (THEORIES)

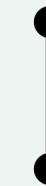


REPRESENTING THE INFORMATION

We would first want to represent the information on a numerical representation. How? Especially for data such as **images, texts, and 1-dimensional signals**



Lorem ipsum dolor sit amet,
consectetur adipiscing elit.
Maecenas mattis fringilla dolor,
vitae dictum ipsum gravida at.
Donec laoreet lacus nec lectus
ullamcorper euismod. Nunc et
erat vestibulum, ...



Features

- Color-based
- Texture-based
- Shape-based



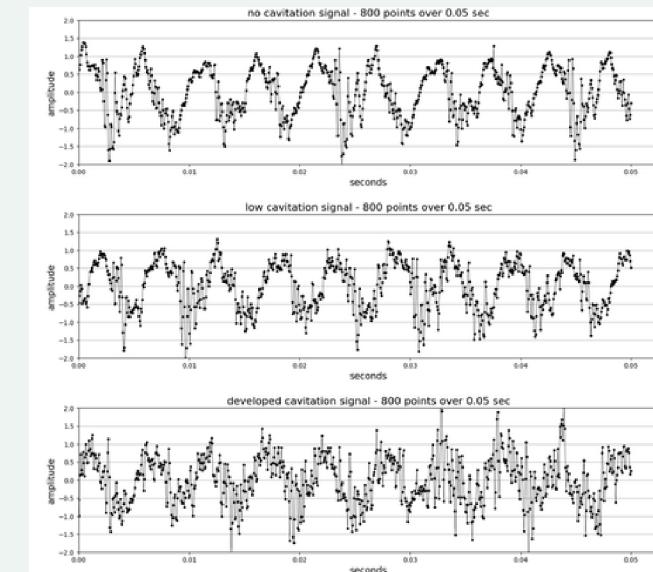
Features

- Frequency-based (Count, Bag of Words, TF-IDF)
- Embedding-based (Vec2Vec, GLOVE)



Features

- Time-domain (Zero-crossing rate, energy)
- Frequency-domain (signal-to-noise)
- Time and Frequency-domain (Spectrogram)



UTILIZING THE INFORMATION/PATTERNS

The information we have just extracted from the unstructured data can be used for prediction in **two ways**.

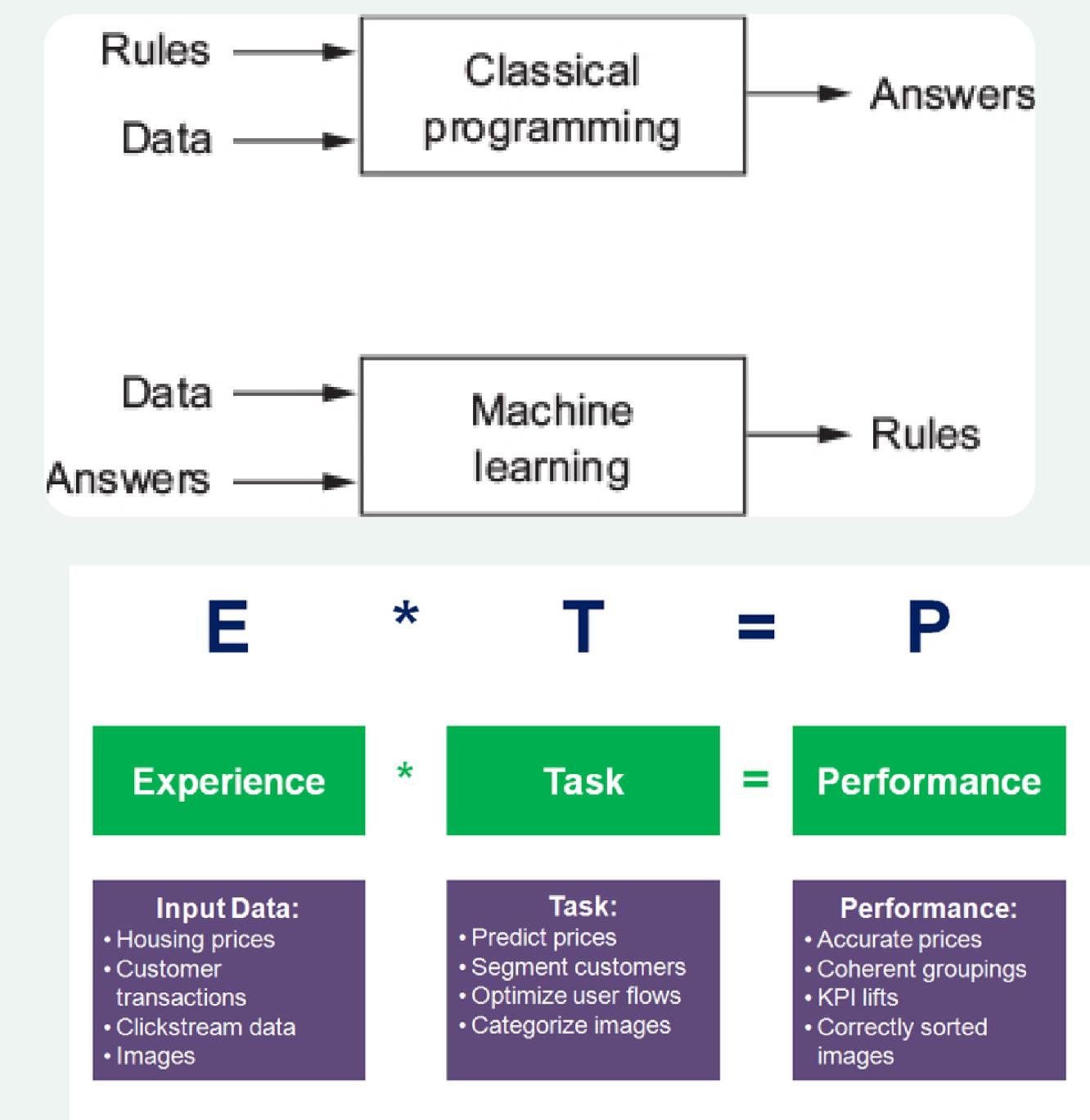
Rule-Based Prediction

We already have a **defined rule**, so that we only input the acquired data to have the prediction result. A prime example is diagnosing a disease based on symptoms.

(Machine) Learning-Based Prediction

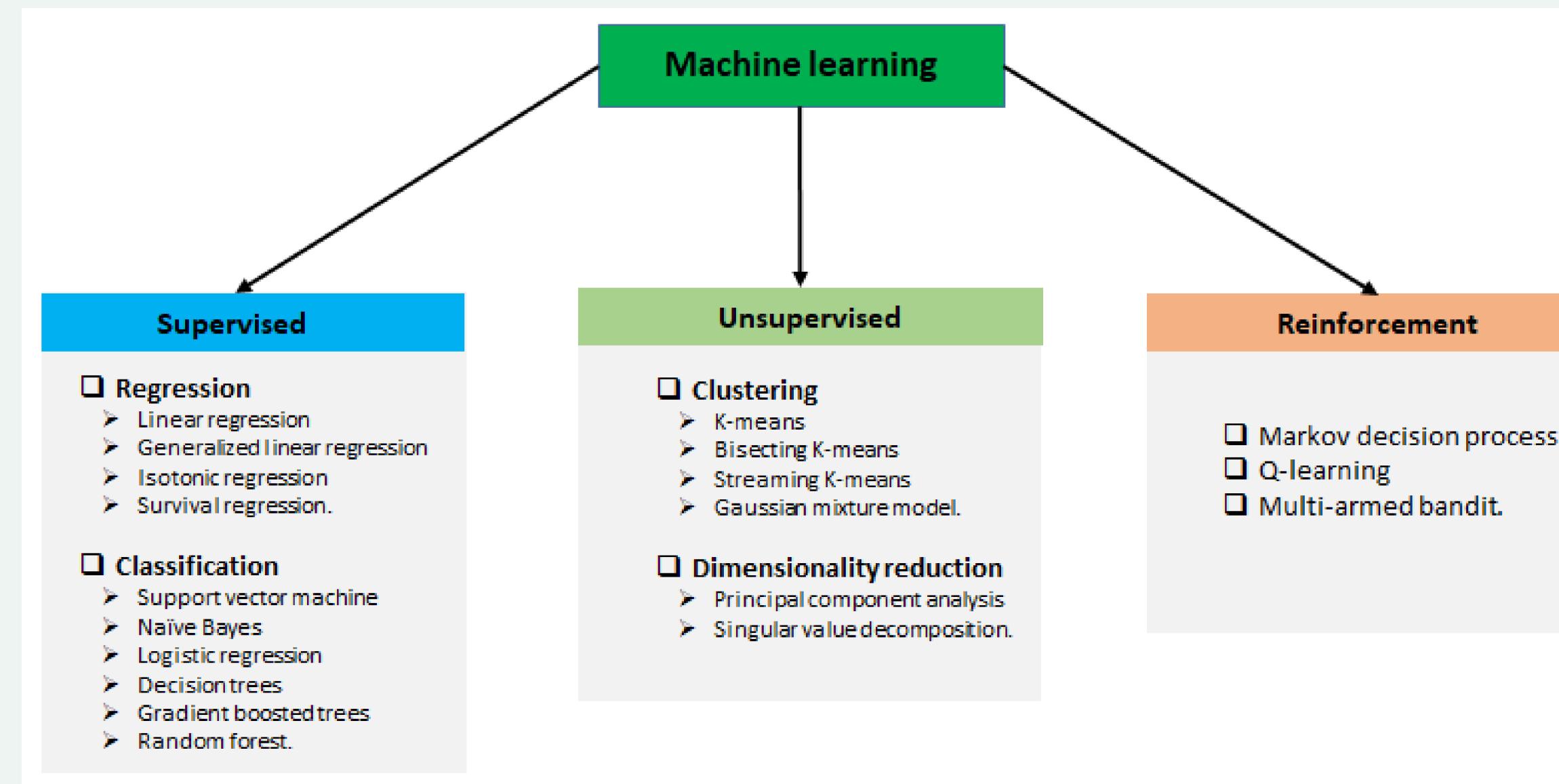
we want do not know **how to determine the output**, so we ask a machine to try to **learn the patterns** in order to make a prediction. This method is a subset of artificial intelligence, called machine learning

By the framework of Experience, Task and Performance (Mitchell, 1997), we can create a machine learning model that can predict, cluster, or re-learn from its own experience.

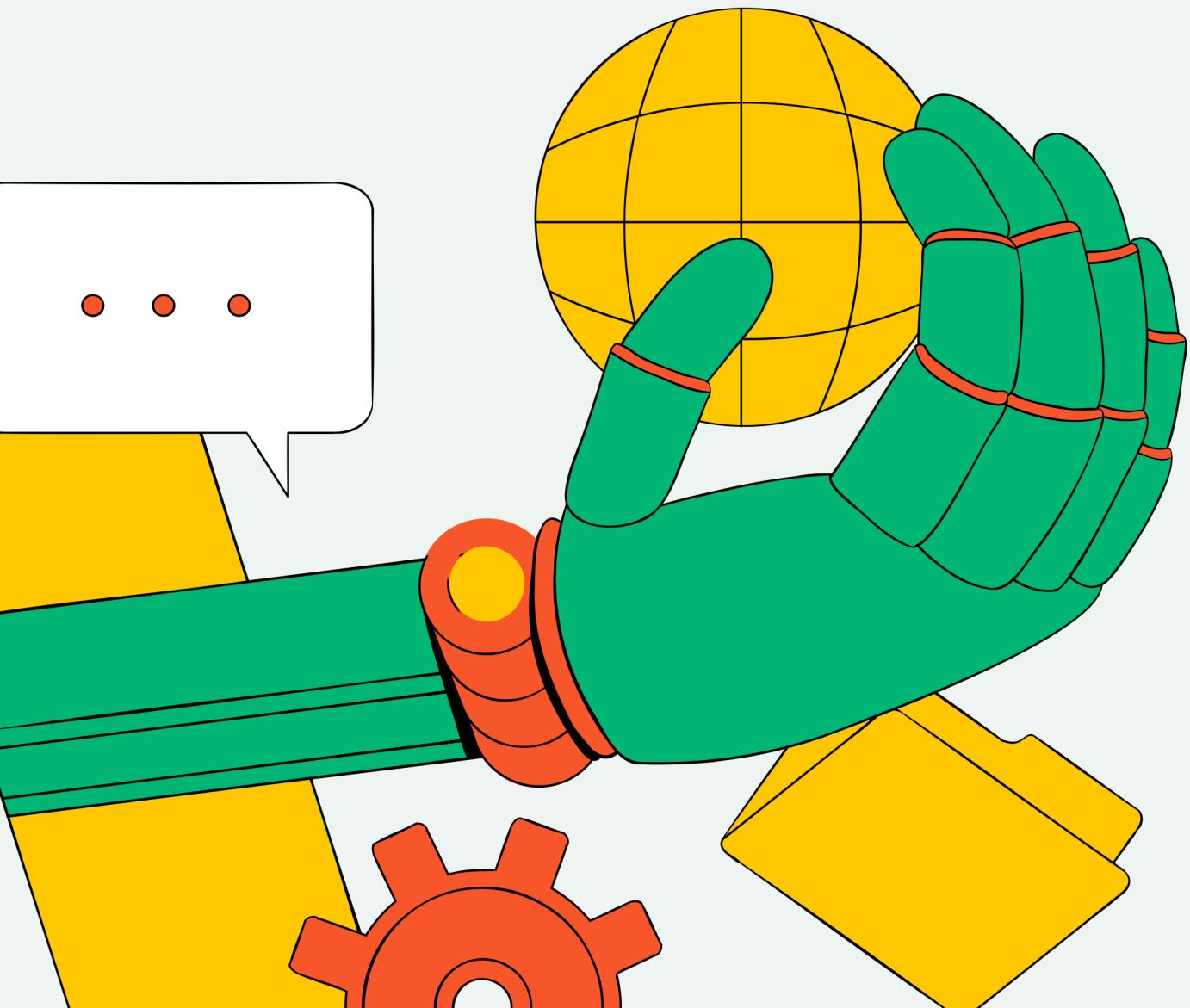


(OTHER) TASK ON MACHINE LEARNING

Aside from supervised learning where given a data of features and target, unsupervised have no target which it could try to cluster or reduce the dimension of the data. Reinforcement learning on the other-hand uses experience as an active learning mechanism.



LET'S TRY HANDS-ON



1. Matos, A., Pinto, B., Barros, F., Martins, S., Martins, J., & Au-Yong Oliveira, M. (03 2019). Smart Cities and Smart Tourism: What Future Do They Bring?
doi:10.1007/978-3-030-16187-3_35
2. Mitchell, T. M. (1997). Machine learning (Vol. 1). McGraw-hill New York.