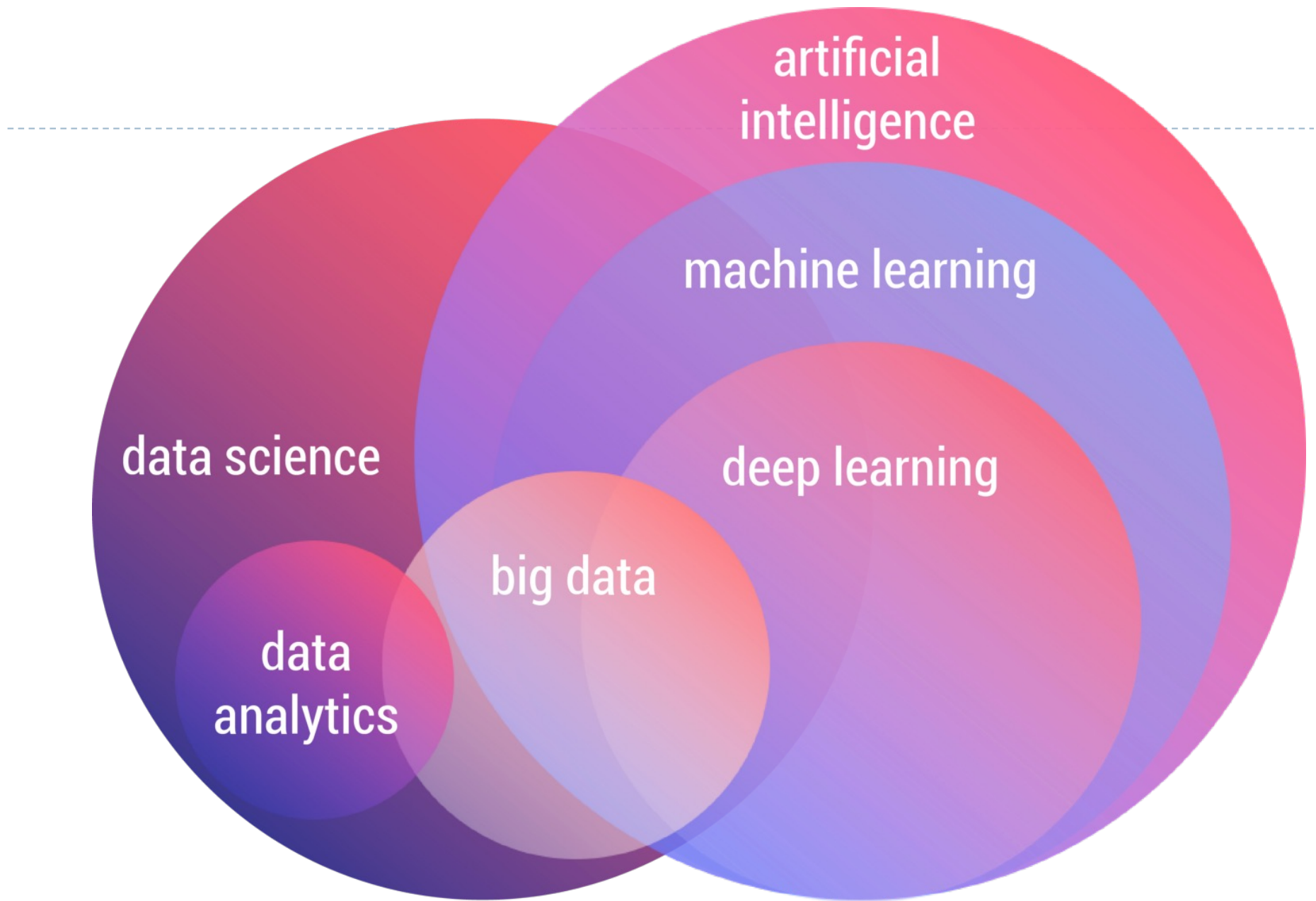




Pelatihan ABCD

Modul 4-1: Machine Learning

Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung
Unviersitas Singaperbangsa Karawang



Contents

- ▶ What is Machine Learning?
 - ▶ How does Machine Learning Work?
 - ▶ Machine Learning Use Cases
 - ▶ Types of Machine Learning
- ▶ Supervised Learning
 - ▶ What is Supervised Learning
 - ▶ Supervised Learning Process
 - ▶ Types of Supervised Learning
 - ▶ Advantages and Disadvantages of Supervised Learning
- ▶ Unsupervised Learning
- ▶ Machine Learning Using Python

What is Machine Learning?

Machine learning is a concept which allows the machine to **learn from examples and experience**, and that too **without being explicitly programmed**

What is Machine Learning?

- ▶ Have you ever shopped online? While checking for a product, did you noticed when it recommends for a product similar to what you are looking for? Or did you noticed “the person bought this product also bought this” combination of products.
- ▶ **How are they doing this recommendation?** This is machine learning



What is Machine Learning?



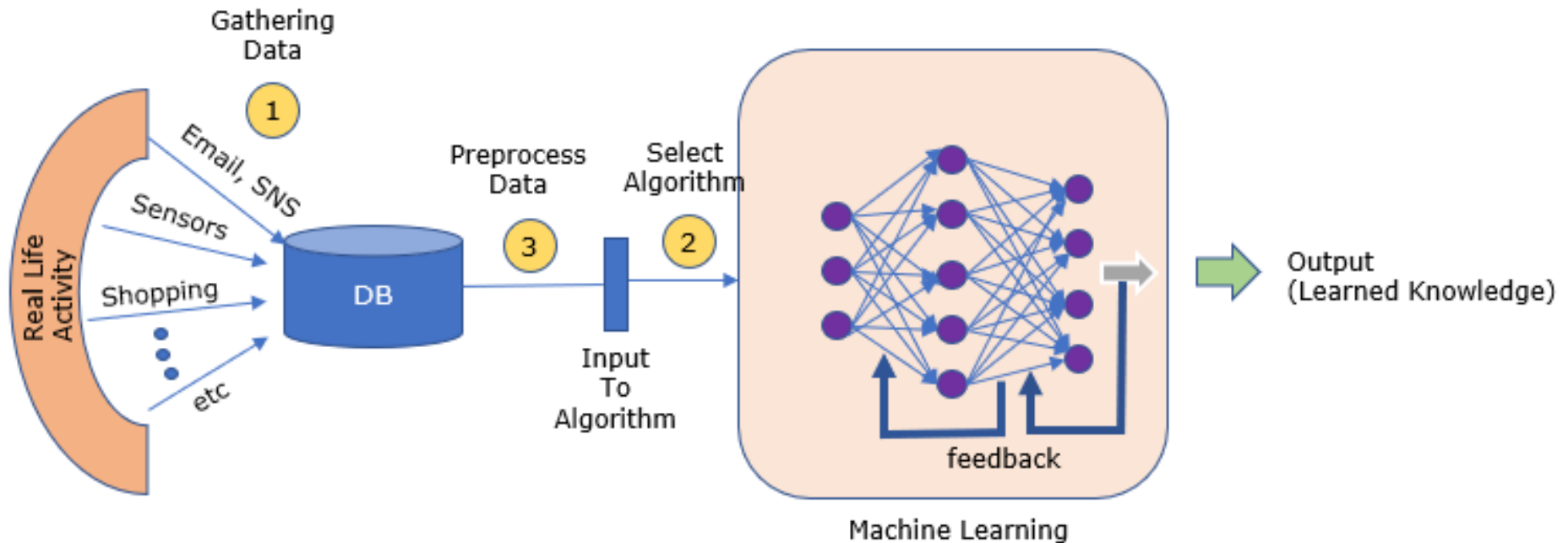
- ▶ Did you ever get a call from any bank or finance company asking you to take a loan or an insurance policy? What do you think, do they call everyone? No, they call only a few selected customers who they think will purchase their product.
- ▶ How do they select? **This is target marketing and can be applied using Clustering.** This is machine learning.

What is Machine Learning?

Machine Learning (ML) is a subset of artificial intelligence which focuses mainly on machine **learning from their experience** and **making predictions based on its experience**

It enables the computers or the machines to **make data-driven decisions** rather than being explicitly programmed for carrying out a certain task. These programs or **algorithms are designed in a way they learn and improve over time when they are exposed to new data**

What is Machine Learning?



- ▶ Machine learning is a method of data analysis that automates analytical model building.
- ▶ Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

How does Machine Learning Work?

- ▶ ML algorithm is **trained using a training data set** to **create a model**. When new input data is introduced to the ML algorithm, it **makes a prediction on the basis of the model**.
- ▶ **The prediction is evaluated for accuracy** and if the accuracy is acceptable, the ML algorithm is deployed. If the accuracy is not acceptable, the ML algorithm is trained again and again with an augmented training data set.

(This is just a brief explanation, as there are many factors and other steps involved)

How does Machine Learning Work?

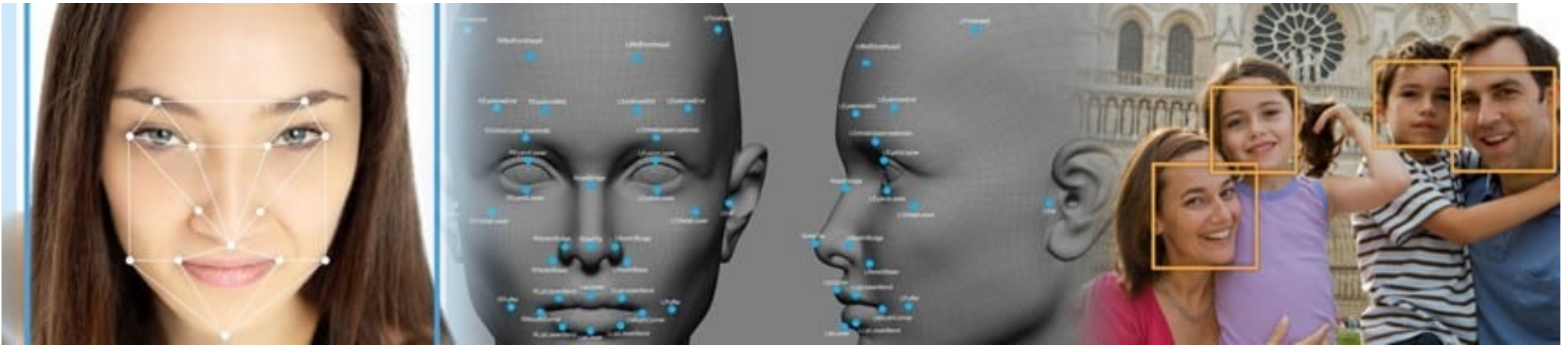
edureka!

What is Machine Learning Use For?

- ▶ Fraud detection
- ▶ Web search results
- ▶ Real-time ads on web pages
- ▶ Credit scoring
- ▶ Prediction of equipment failures
- ▶ New pricing models
- ▶ Network intrusion detection
- ▶ Recommendation engines
- ▶ Customer segmentation
- ▶ Text sentiment analysis
- ▶ Customer churn
- ▶ Pattern and image recognition
- ▶ Email spam filtering

Machine Learning Use Case

► Face Detection

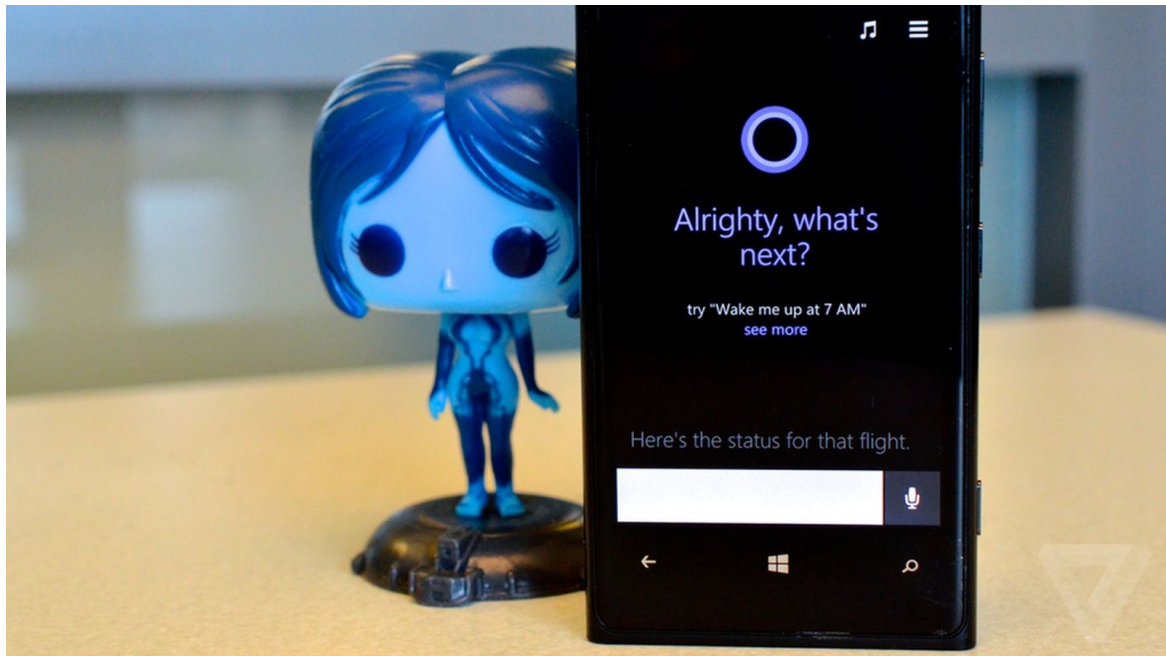


► Mask Detection



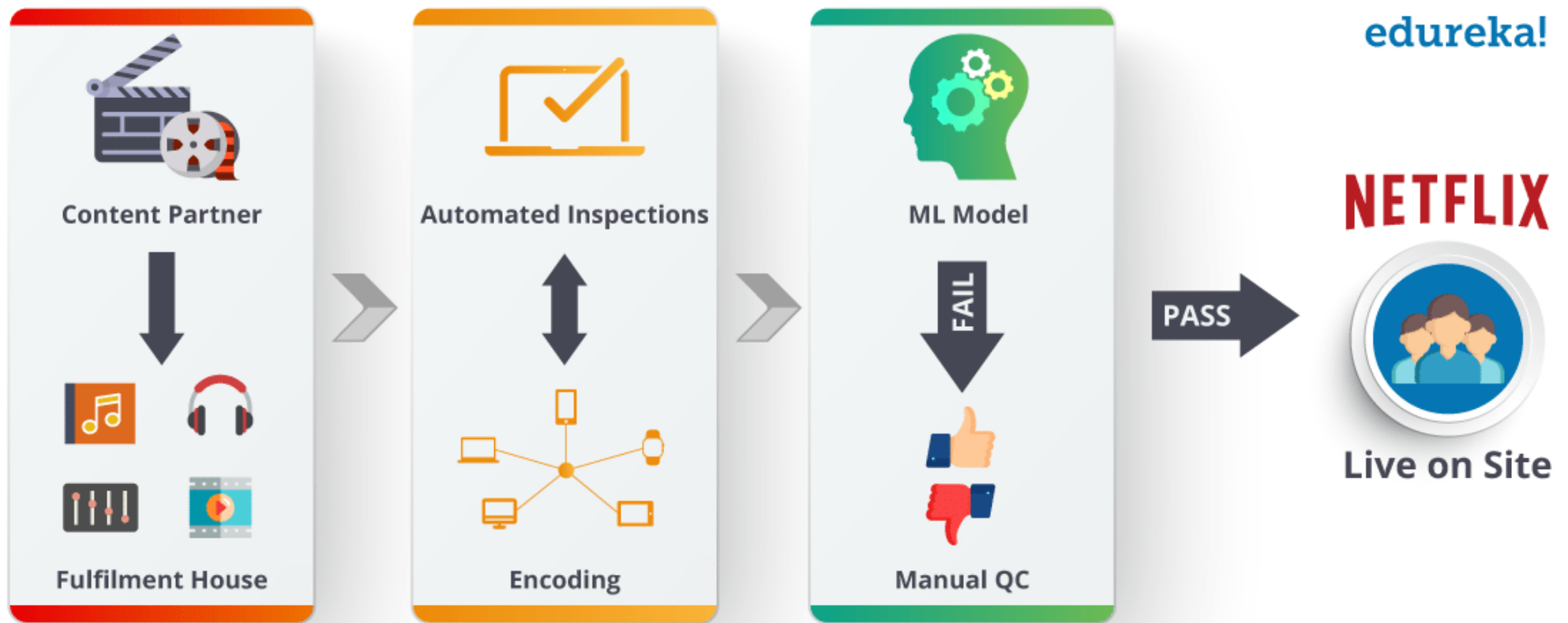
Machine Learning Use Case

- ▶ **Cortana**, virtual assistant developed by Microsoft (2014) which uses the Bing search engine to perform tasks such as setting reminders and answering questions for the user.
- ▶ Speech automated system. Train voice and start working based on this training.

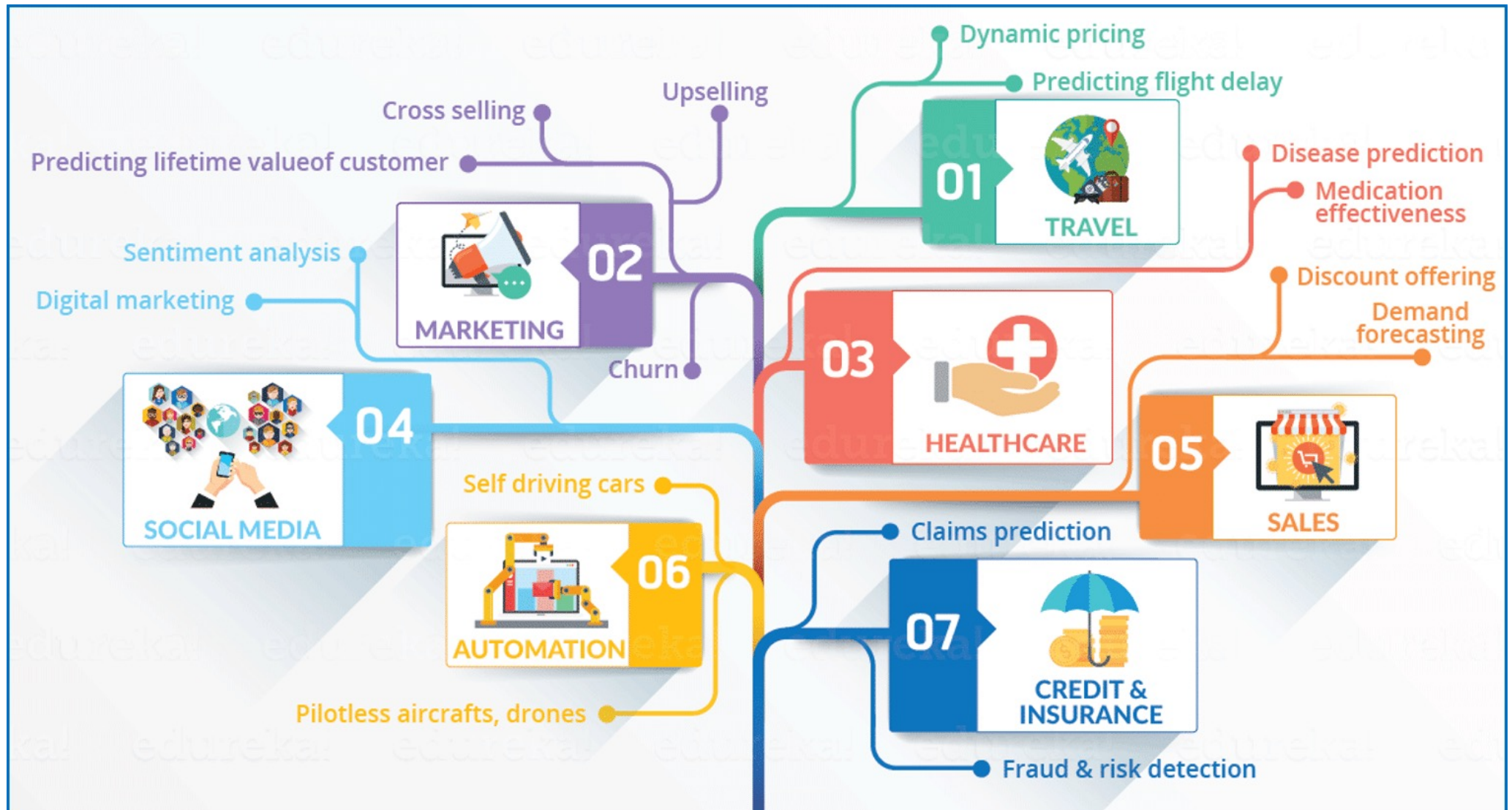


Machine Learning Use Case

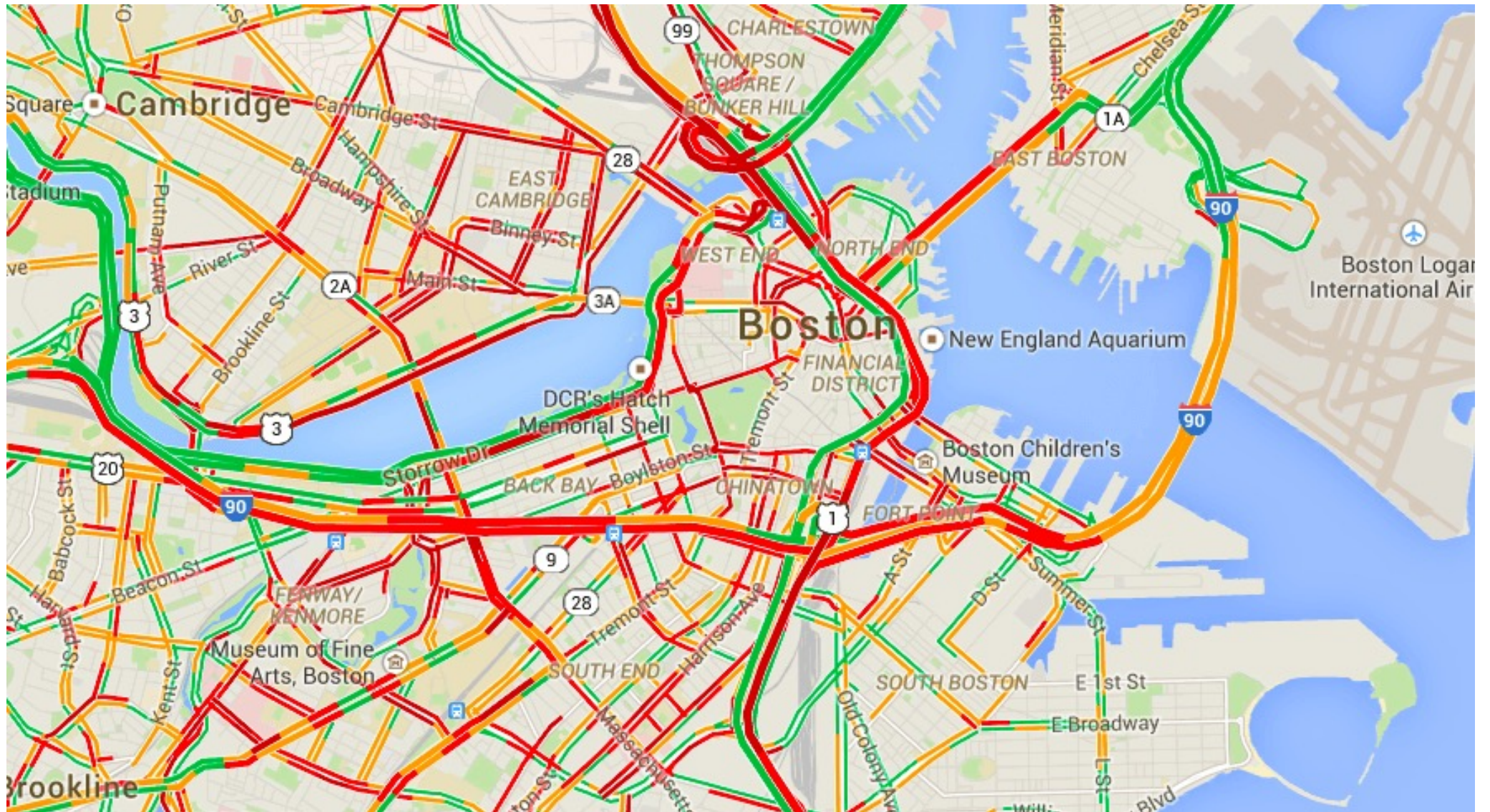
► Recommendation System



Machine Learning Use Case: Prediction



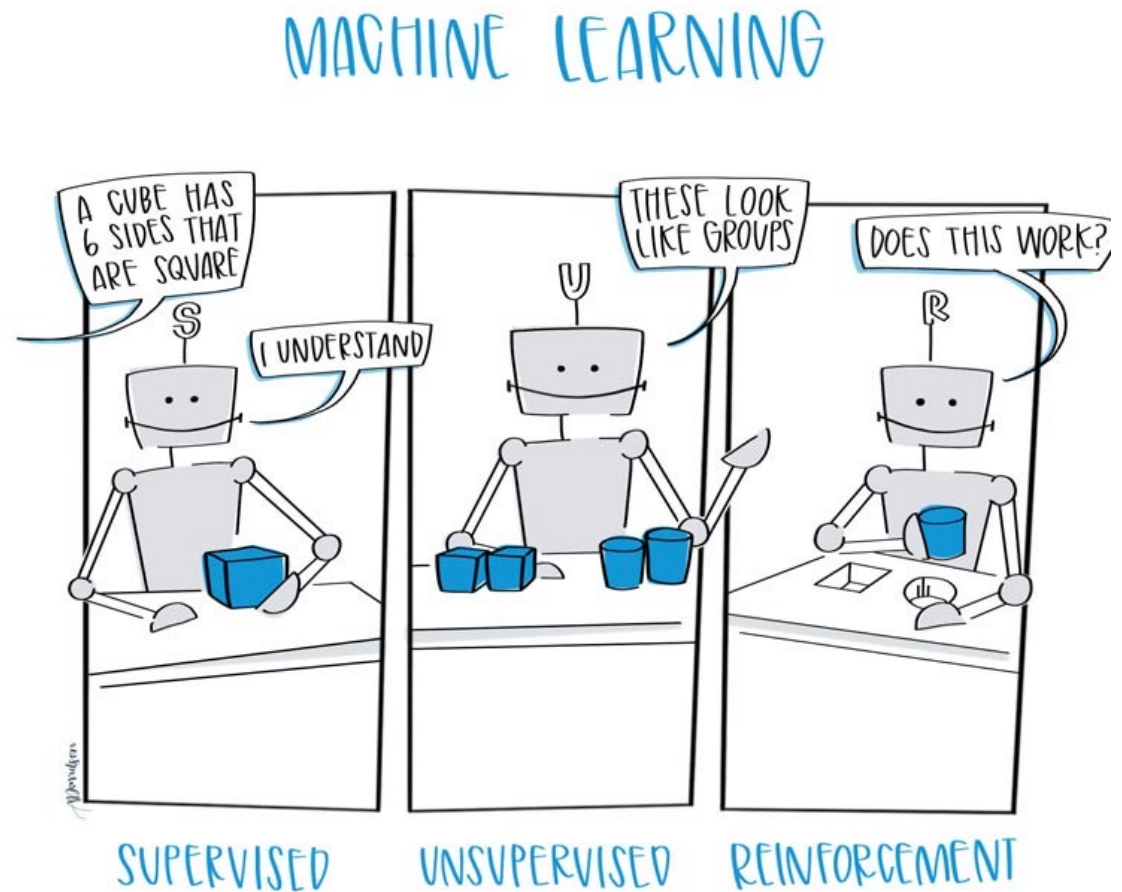
Google Map Traffic Analysis



Types of Machine Learning

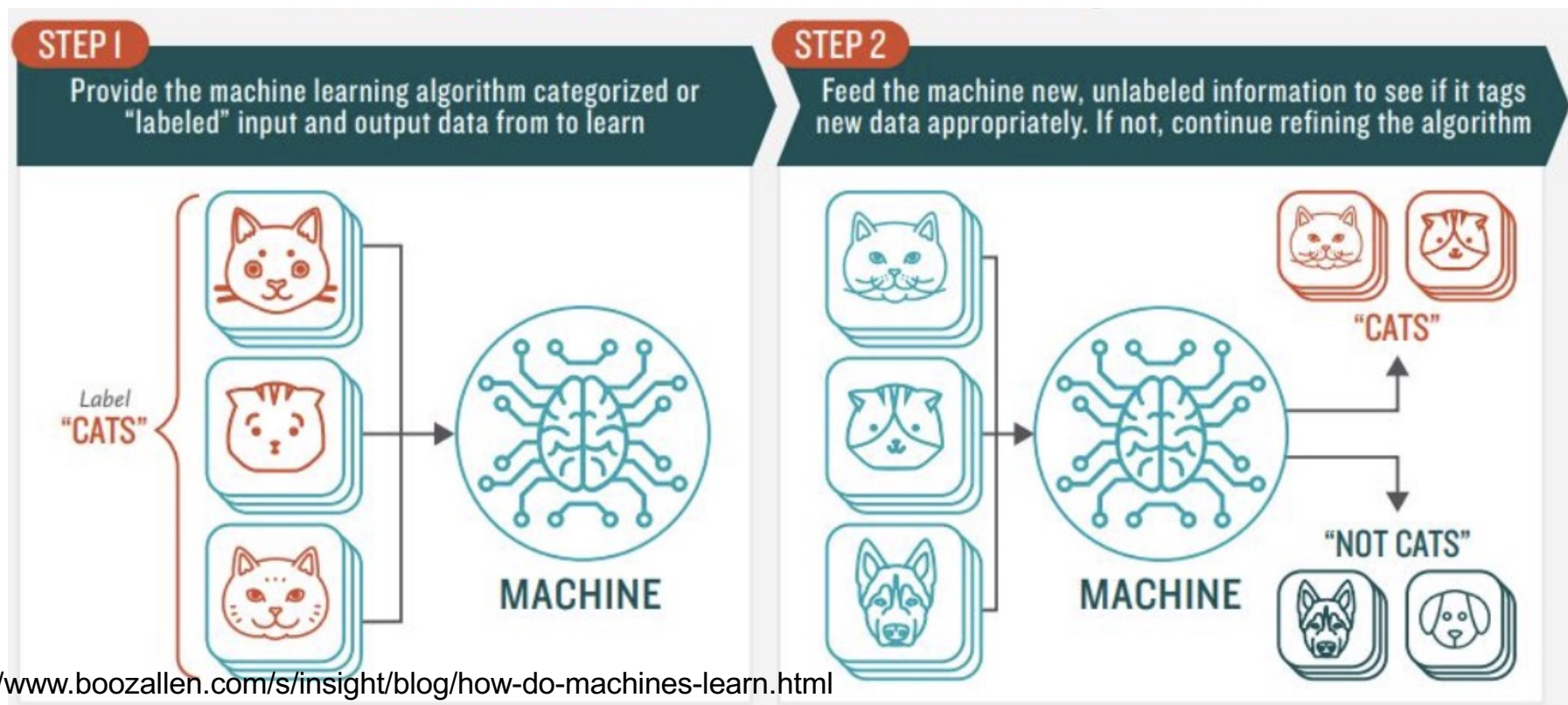
ML is sub-categorized to three types:

1. **Supervised Learning** – Train Me!
2. **Unsupervised Learning** – I am self sufficient in learning
3. **Reinforcement Learning** – My life My Rules! (Hit & Trial)



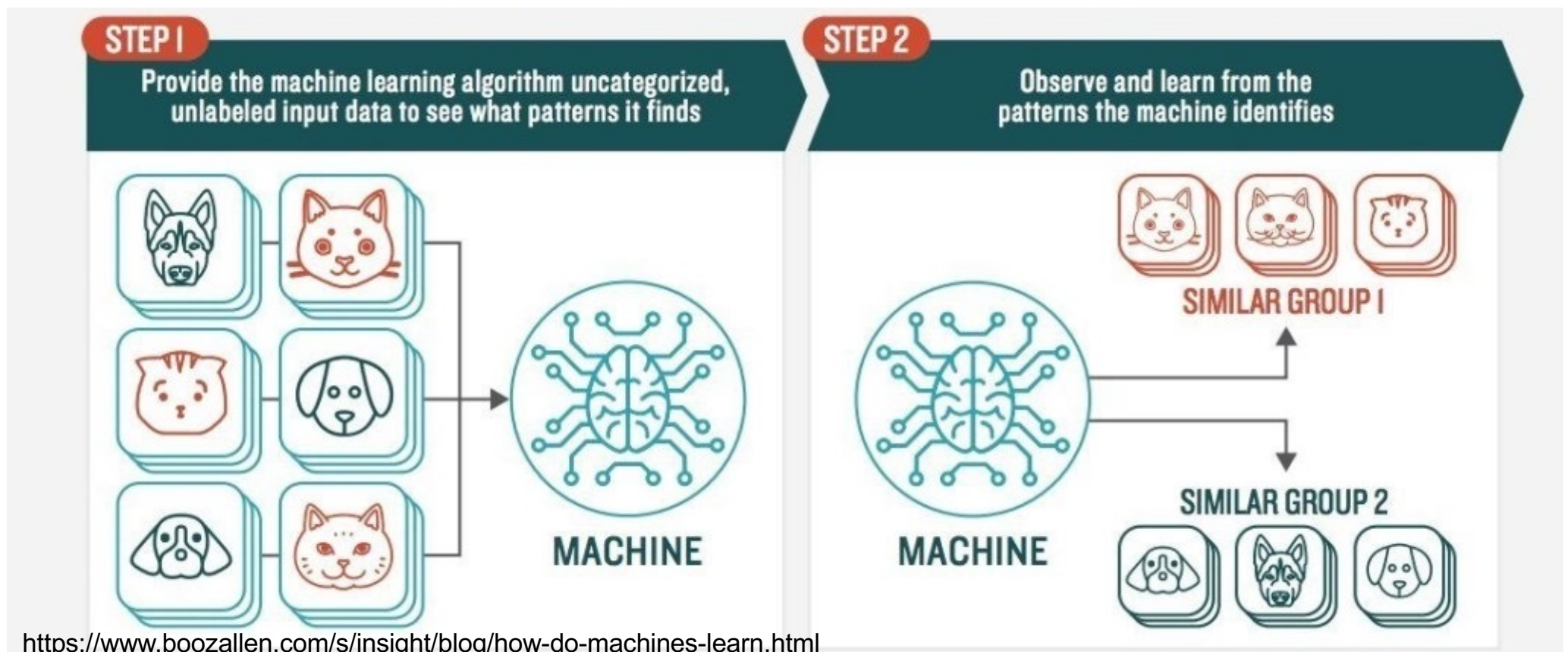
What is Supervised Learning?

- ▶ Supervised Learning is the one, where you can consider **the learning is guided by a teacher**.
- ▶ We **have a dataset** which acts as a teacher and its role is to train the model or the machine.
- ▶ Once the model gets trained it can start making a prediction or decision when new data is given to it.



What is Unsupervised Learning?

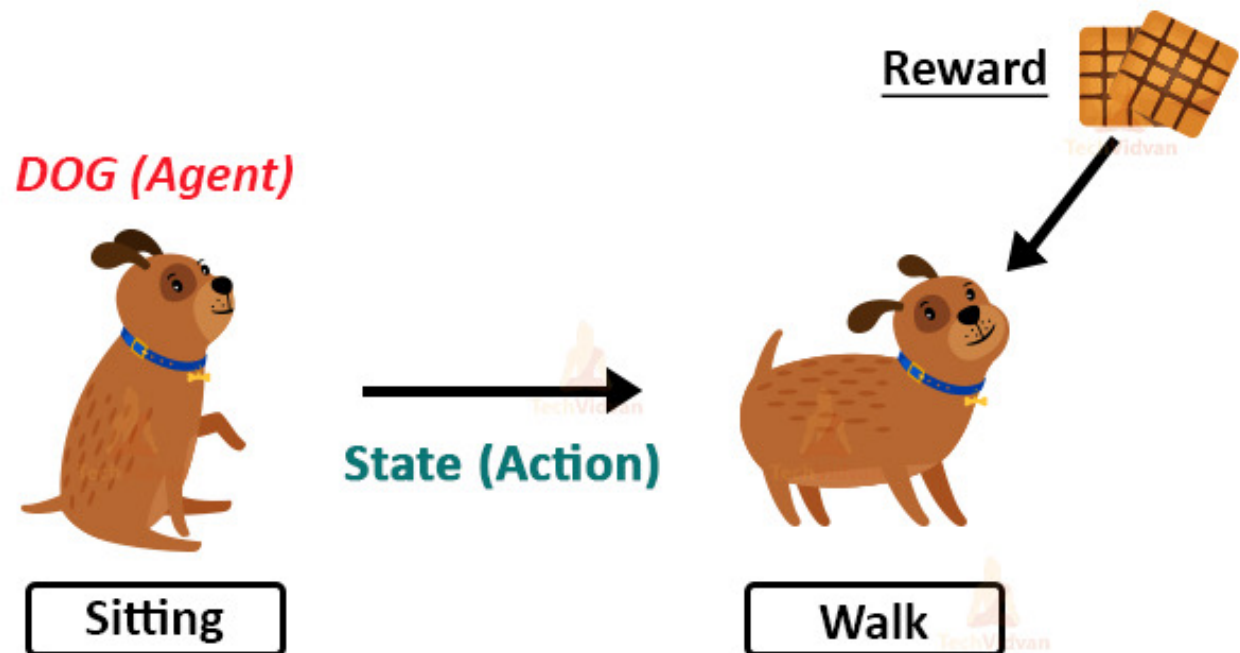
- ▶ The model **learns through observation** and **finds structures in the data**.
- ▶ Once the model is given a dataset, it **automatically finds patterns, relationships, feature similarity in the dataset** by **creating clusters in it**.
- ▶ **What it cannot do** is add **labels to the cluster**. For example, it cannot say this group is cats or dogs, but it will separate all to the similar group.



What is Reinforcement Learning?

- ▶ Reinforcement Learning is **enforcing models to learn how to make decisions**.
- ▶ It is the ability of an agent to **interact with the environment and find out what is the best outcome**. It follows the concept of **hit and trial method**.
- ▶ The agent is **rewarded or penalized** with a point for a correct or a wrong answer, and on the basis of the positive reward points gained the model trains itself. And again once trained it gets ready to predict the new data presented to it.

- To train your dog to sit, there is certain instructions to the dog to try to make it learn.
- If the dog executes the instruction perfectly, it would get a biscuit as a reward. If not, it would not get anything.



Types of Machine Learning

Supervised

- Data has known labels or output

- Insurance underwriting
- Fraud detection
- Spam detection
- Image classification
- Risk assessment

Unsupervised

- Labels or output unknown
- Focus on finding patterns and gaining insight from the data

- Customer clustering
- Text mining
- Face recognition
- Big data visualization
- Targetted marketing

Reinforcement

- Focus on making decisions based on previous experience
- Policy making with feedback

- Gaming
- Finance sector
- Reward systems
- Robot navigation
- Self-driving car

Need to Explore – Characteristic and Opportunities

Supervised

- May get a good results

- Manual labelling is usually slow, costly and error prone

Unsupervised

- No need manual labelling/ annotation
- Affordable way of data exploration

- It may take (much) longer to get the accurate model
- We have little control over what patterns it will find.

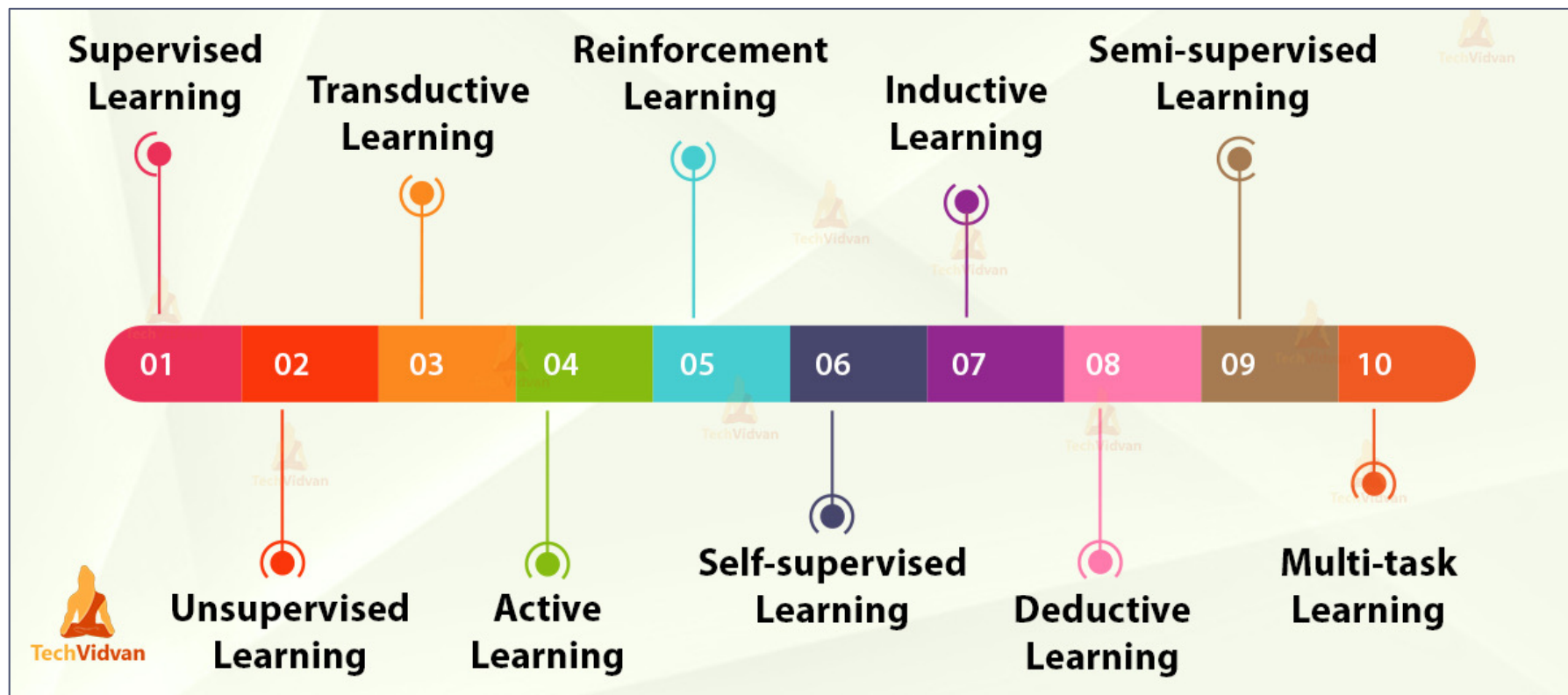
Reinforcement

- Need no manual labelling of training data
- May produce a more controlled output

- Need to explain to the 'machine' what is 'win' and what is 'fail'.
- The machine learn sequence of actions needed to 'win'

Types of Machine Learning - Expanded

- ▶ Machine Learning is a very vast subject and every individual field in ML is an area of research in itself.
- ▶ The subject is expanding at a rapid rate due to new areas of studies constantly coming forward.



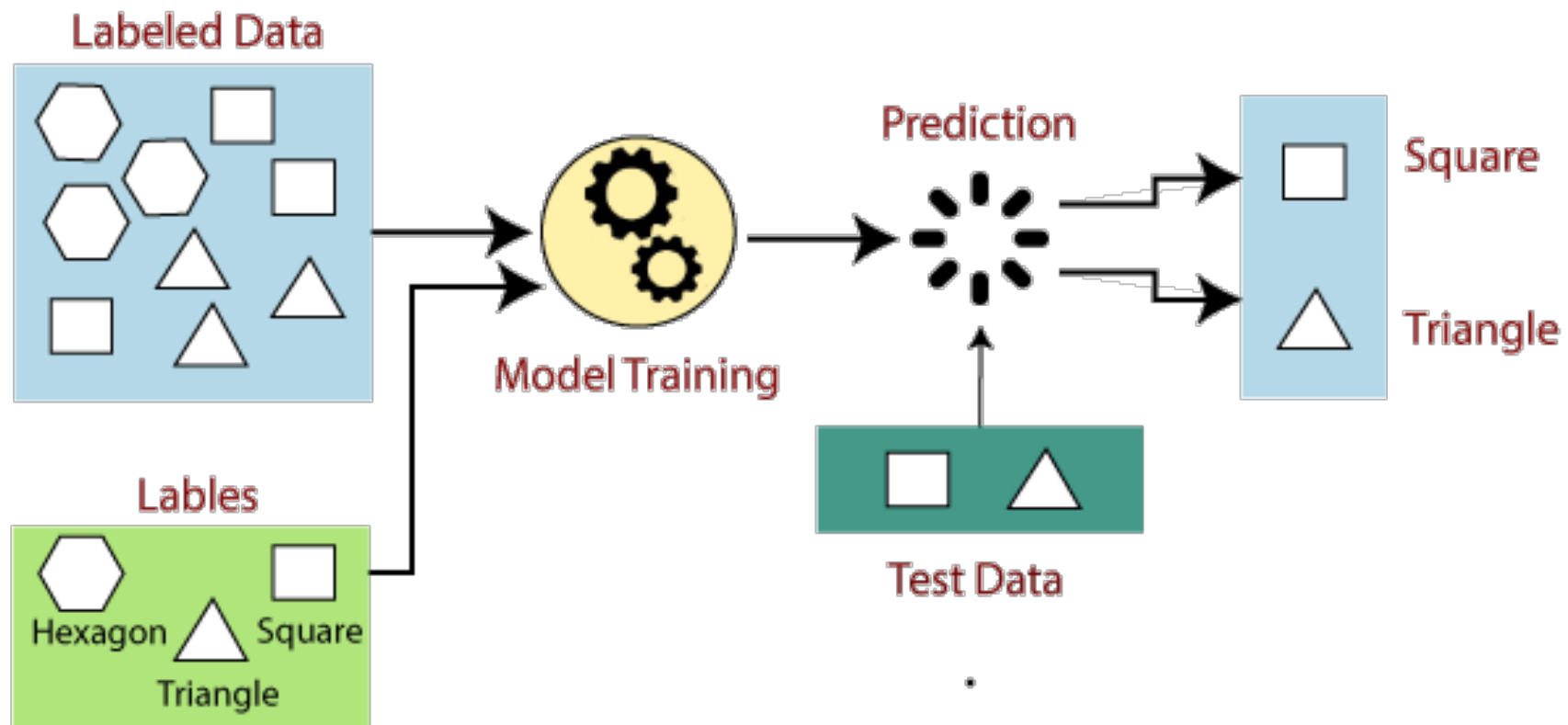
Supervised Learning

Supervised Learning

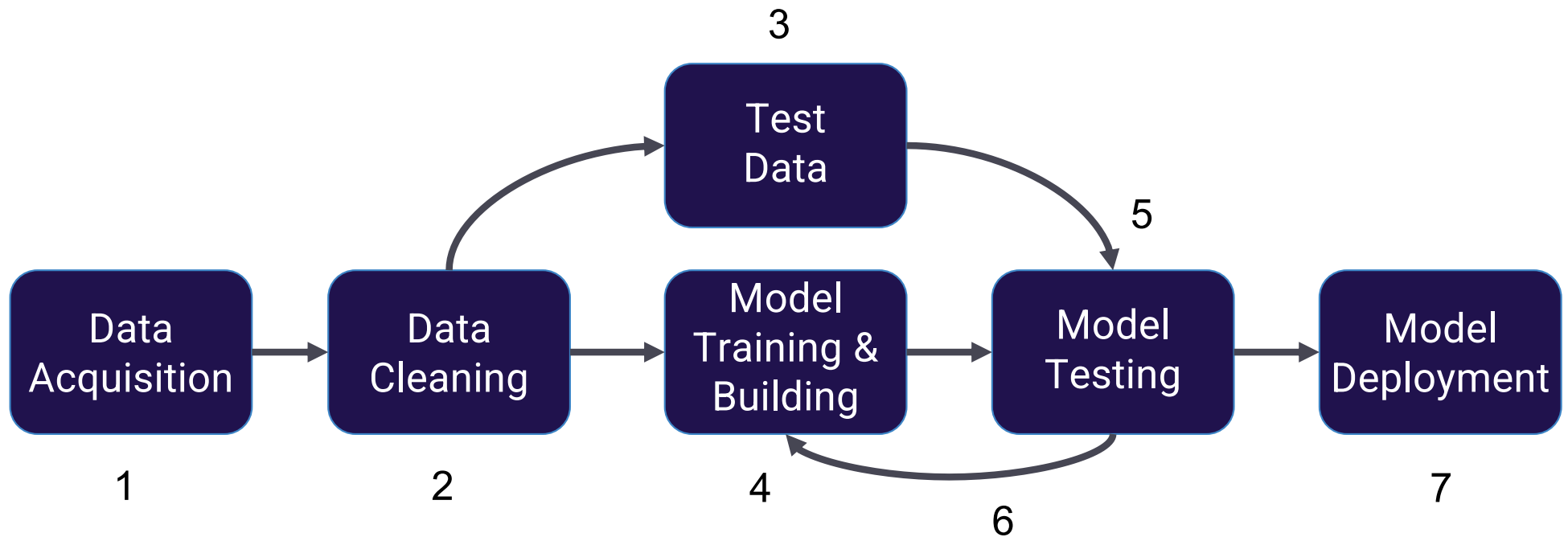
- ▶ Types of machine learning in which machines are **trained using well "labelled" training data**, and on basis of that data, **machines predict the output**.
- ▶ The labelled data means some input data is already tagged with the correct output.
- ▶ The training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the **same concept as a student learns in the supervision of the teacher**.
- ▶ Supervised learning is a process of providing input data as well as correct output data to the machine learning model. The aim of a supervised learning algorithm is to **find a mapping function to map the input variable(x) with the output variable(y)**.

Supervised Learning

- Models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output.



Supervised Learning Process



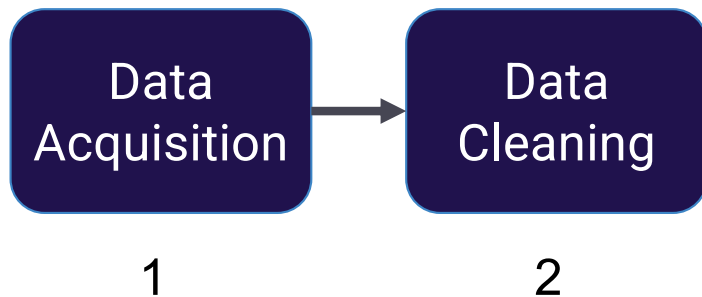
Process 1: Data Acquisition

- ▶ Get the data
 - ▶ Data of customers, collecting from data base online, data from sensors, etc.

Data
Acquisition

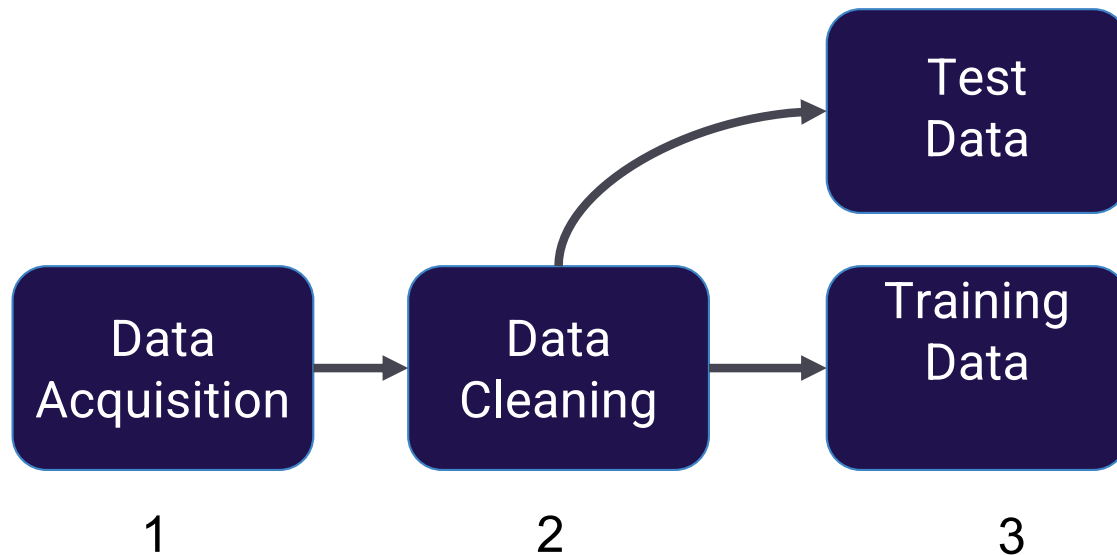
Process 2: Data Cleaning

- ▶ Cleaning and formating data
 - ▶ using library, e.g. Pandas
 - ▶ Library like Dask and Vaex for big data



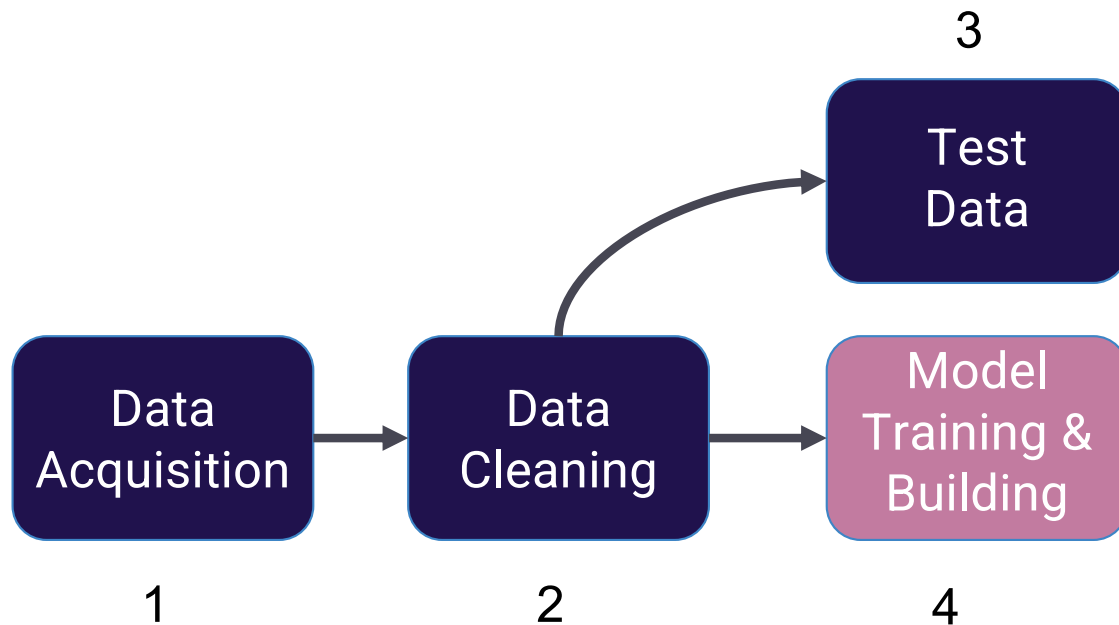
Process 3: Split the Data

- ▶ Split the data into Training Data and Test Data
 - ▶ e.g. 30% for test data and 70% for training data



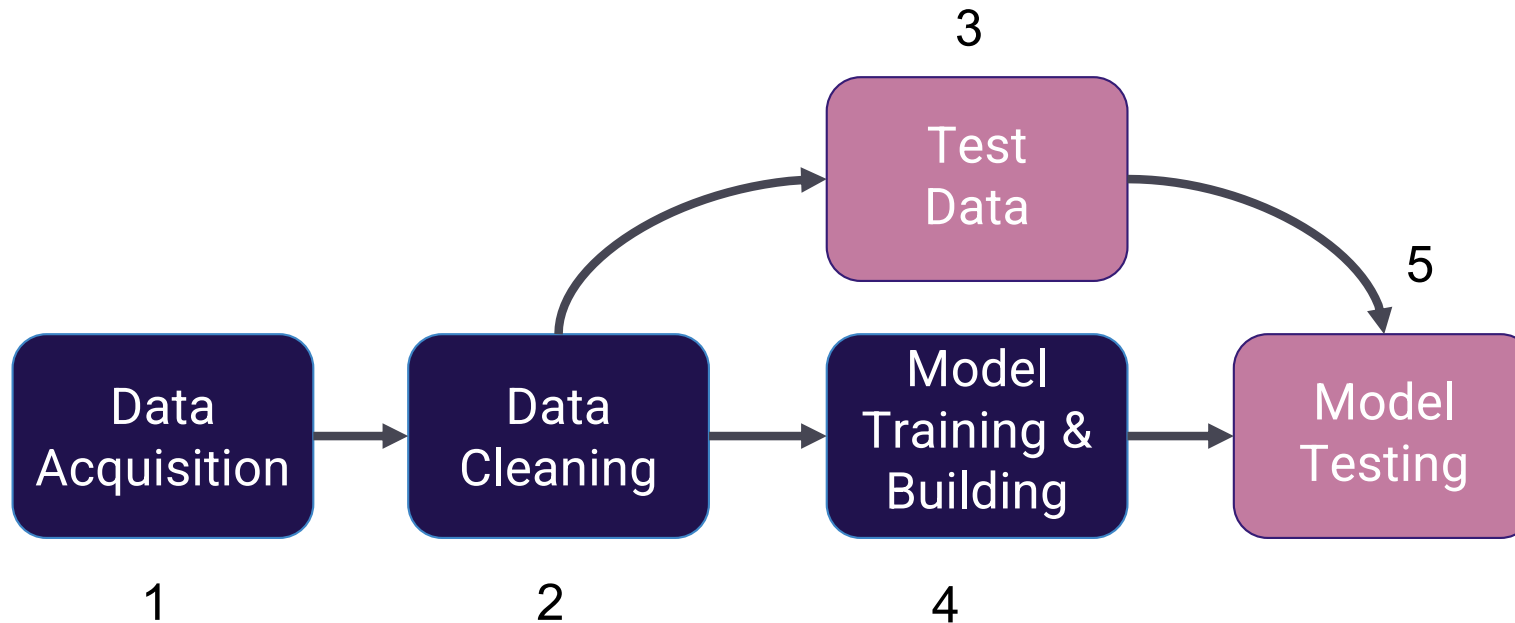
Process 4: Model Training

- ▶ Train model using training data



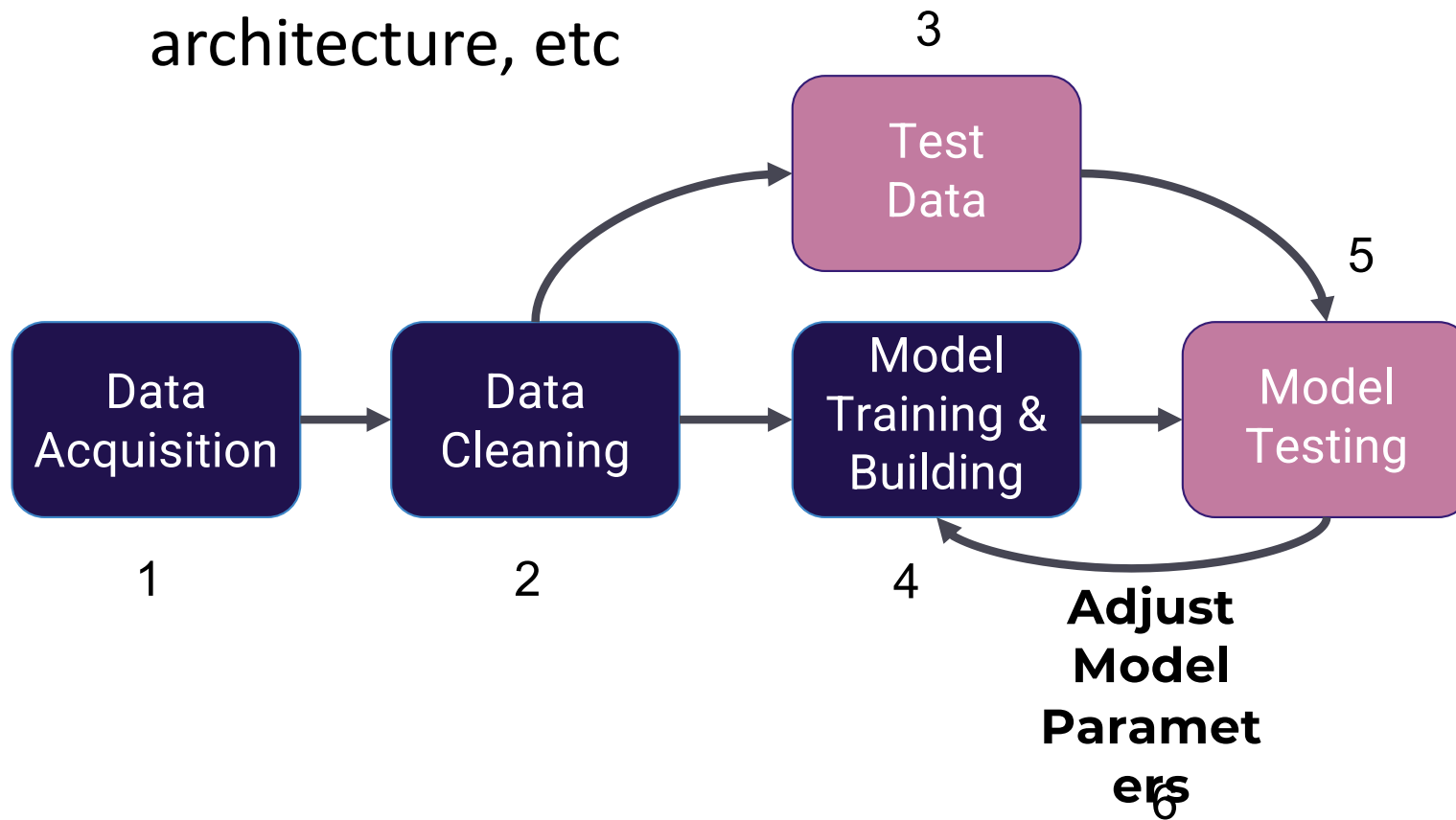
Process 5: Model Performance Testing

- ▶ Run test data through the model. Compare the model prediction to the actual correct label from the test data



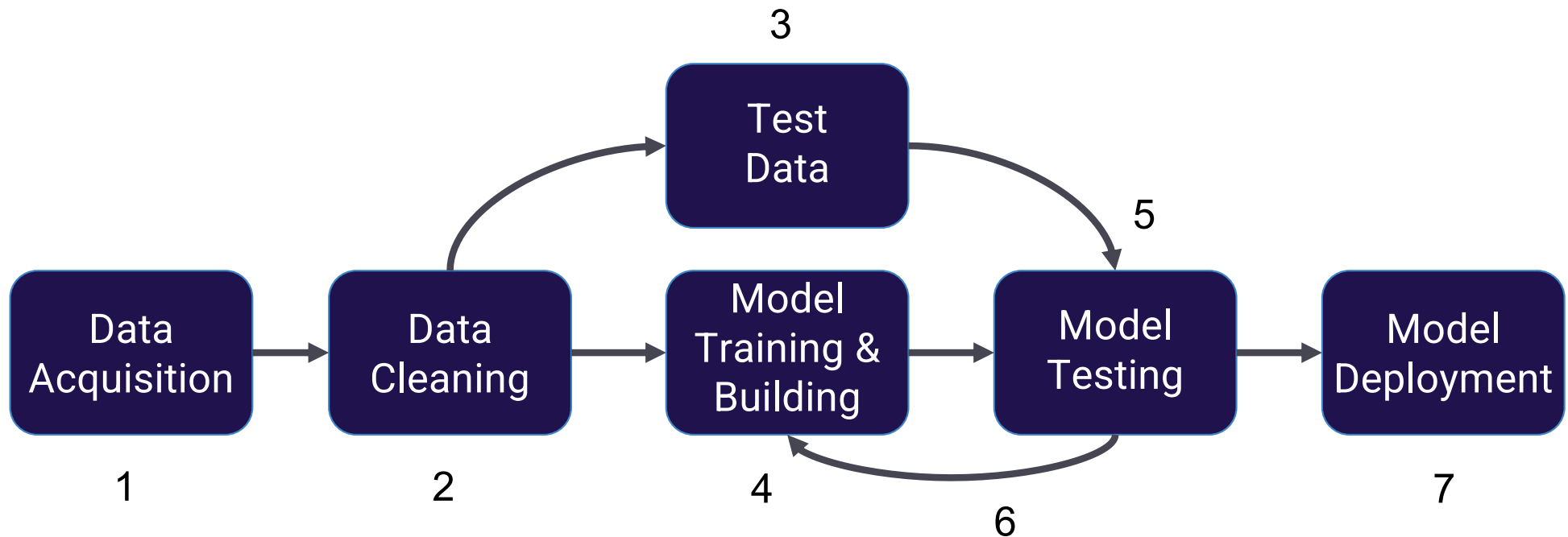
Process 6: Adjust Model Parameters

- Evaluate the model. If needed go back to adjust model parameters: add more neurons, add more layers, change architecture, etc



Process 7: Deploy the Model

- ▶ If the model already meet performance requirement, deploy the model to the real world



Supervised Learning Data Set

- ▶ In Simplified approach of supervised learning, we use single split of the data to evaluate our model's performance (only test data).
- ▶ **After all, we were given the chance to update the model parameters again and again. This became an issue.**
- ▶ To fix this issue, data is often split into **3 sets**
 1. Training Data
 - ▶ Used to train model parameters
 2. Validation Data
 - ▶ Used to determine what model hyperparameters to adjust
 3. Test Data
 - ▶ Used to get some final performance metric
- ▶ This means **after we see the results on the final test set we don't get to go back and adjust any model parameters!**
- ▶ This final measure is what we label the true performance of the model to be.

Supervised Learning Process - Recap

1. First Determine the type of training dataset
2. Collect/Gather the labelled training data, and then clean the data.
3. Split the training dataset into training **dataset, test dataset, and validation dataset.**
4. Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
5. Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
6. Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
7. Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.

Types of Supervised Learning

Supervised learning can be further divided into two types of problems:

1. Regression

- ▶ Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc.
- ▶ Some popular Regression algorithms which come under supervised learning: Linear Regression, Regression Trees, Non-Linear Regression, Bayesian Linear Regression, Polynomial Regression

2. Classification

- ▶ Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc, for example Spam Filtering.
- ▶ Some popular classification algorithms: KNN (K-Nearest Neighbors), Random Forest, Decision Trees, Logistic Regression, Support vector Machines.

Supervised Learning

Advantages of Supervised Learning:

- ▶ The model can predict the output on the basis of prior experiences.
- ▶ We can have an exact idea about the classes of objects.
- ▶ Supervised learning model helps to solve various real-world problems such as **fraud detection, spam filtering**, etc.

Disadvantages of supervised learning:

- ▶ Not suitable for handling the complex tasks.
- ▶ Cannot predict the correct output if the test data is different from the training dataset.
- ▶ Training required lots of computation times.
- ▶ We need enough knowledge about the classes of object.

Unsupervised Learning

Unsupervised Learning

Supervised learning: where the **label was known** due to **historical labelled data**.

But what happens when we don't have historical labels?

Unsupervised Learning

- ▶ Unsupervised learning is a machine learning technique in which models are not supervised using training dataset. Instead, models itself find the hidden patterns and insights from the given data.
- ▶ It can be compared to learning which takes place in the human brain while learning new things.
- ▶ Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data.
- ▶ The goal of unsupervised learning is to **find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.**

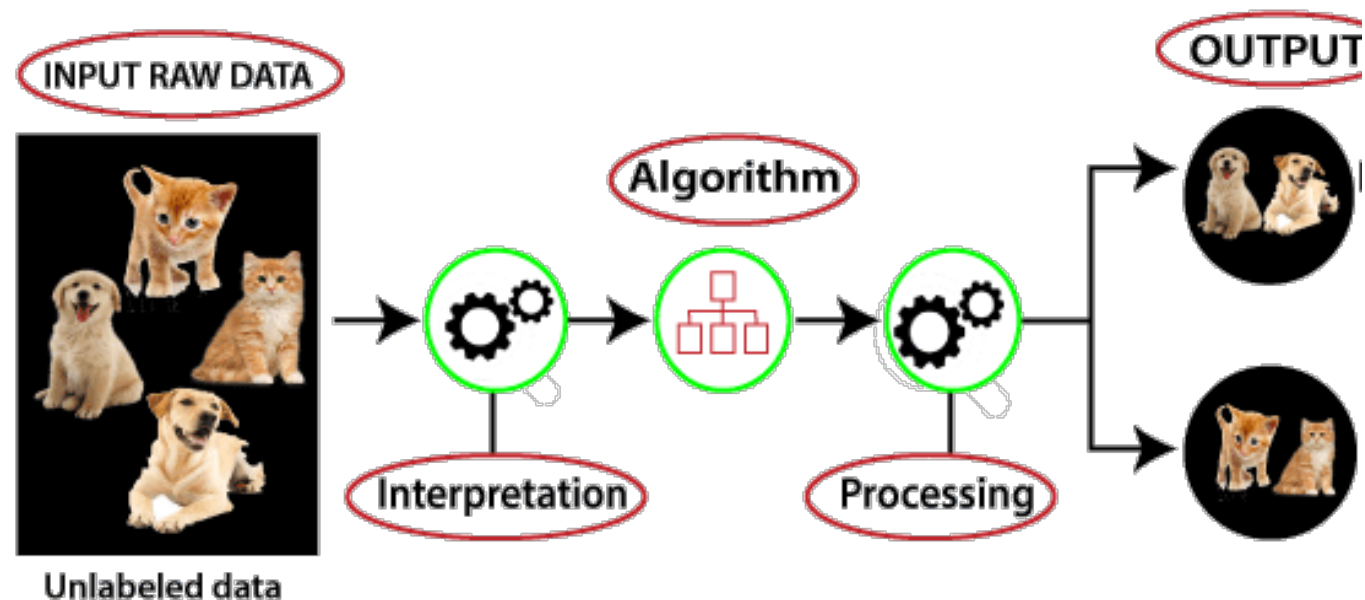
Why Use Supervised Learning

Some main reasons which describe the importance of Unsupervised Learning:

- ▶ Unsupervised learning is helpful for **finding useful insights from the data**.
- ▶ Unsupervised learning is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
- ▶ Unsupervised learning works on unlabeled and uncategorized data which make unsupervised learning more important.
- ▶ In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.

Unsupervised Learning

- ▶ Here, we have taken an unlabeled input data, which means it is not categorized and corresponding outputs are also not given. Now, this unlabeled input data is fed to the machine learning model in order to train it. Firstly, it **will interpret the raw data to find the hidden patterns from the data** and then will apply suitable algorithms.
- ▶ Once it applies the suitable algorithm, the algorithm divides the data objects into groups according to the similarities and difference between the objects.



Type of Unsupervised Learning

There are certain type of problems that fall under unsupervised learning:

- ▶ **Clustering**
 - ▶ **Grouping together unlabeled** data points into categories/clusters
 - ▶ Data points are **assigned to a cluster based on similarity**
 - ▶ Cluster analysis finds the commonalities between the data points/ data objects and categorizes them as per the presence and absence of those commonalities.
- ▶ **Anomaly Detection**
 - ▶ Attempts to **detect outliers** in a dataset
 - ▶ For example, fraudulent transactions on a credit card.
- ▶ **Dimensionality Reduction**
 - ▶ Data processing techniques that reduces the number of features in a data set, either for compression, or to better understand underlying trends within a data set.

Unsupervised Learning Process



Unsupervised Learning Algorithm

Some popular unsupervised learning algorithms:

- K-means clustering
- Hierarchical clustering
- Anomaly detection
- Unsupervised Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

Unsupervised Learning

Advantages of Unsupervised Learning

- Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.
- Unsupervised learning is preferable as it is easy to get unlabeled data in comparison to labeled data.

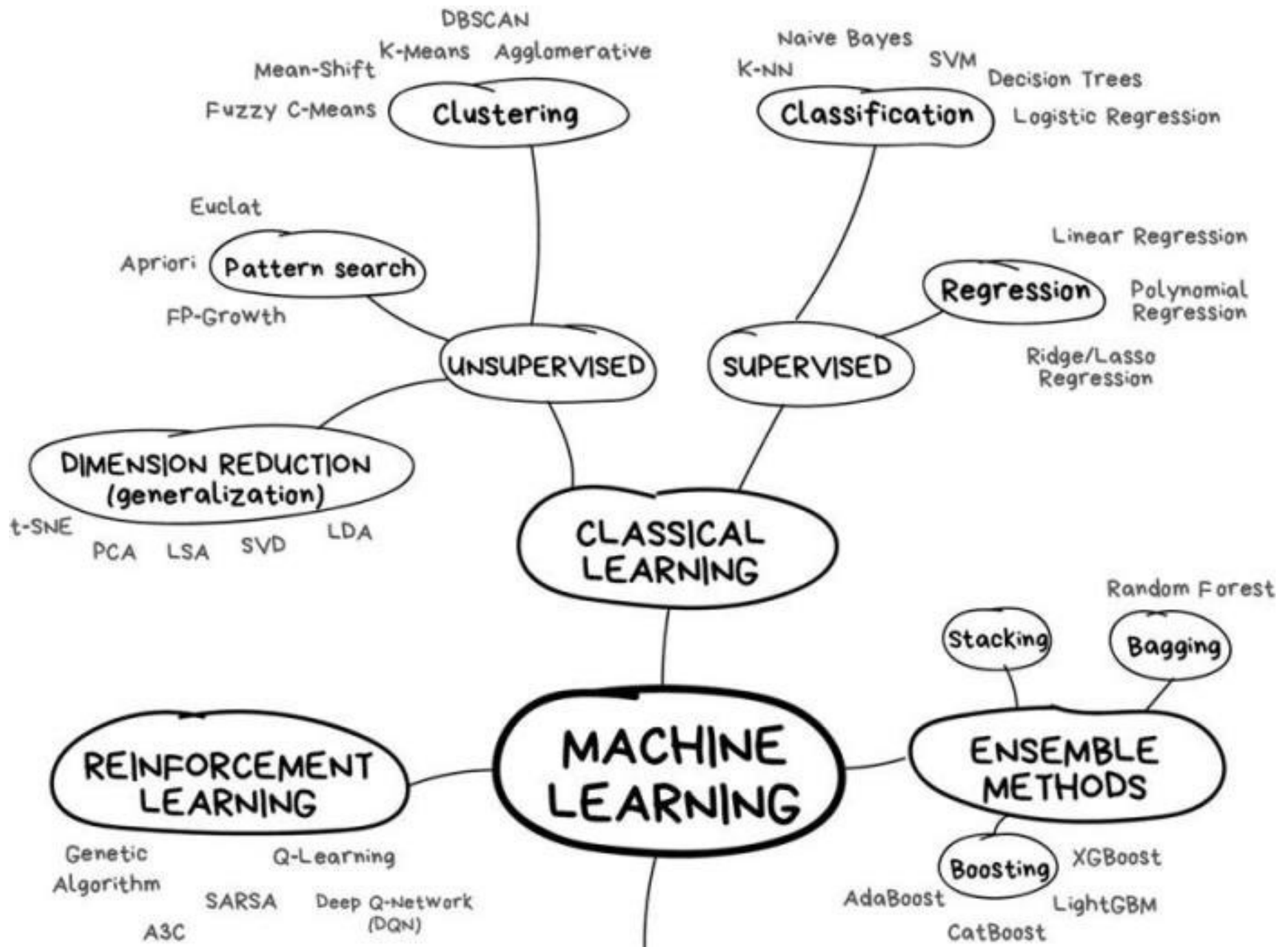
Disadvantages of Unsupervised Learning

- Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
- The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

Machine Learning With Python

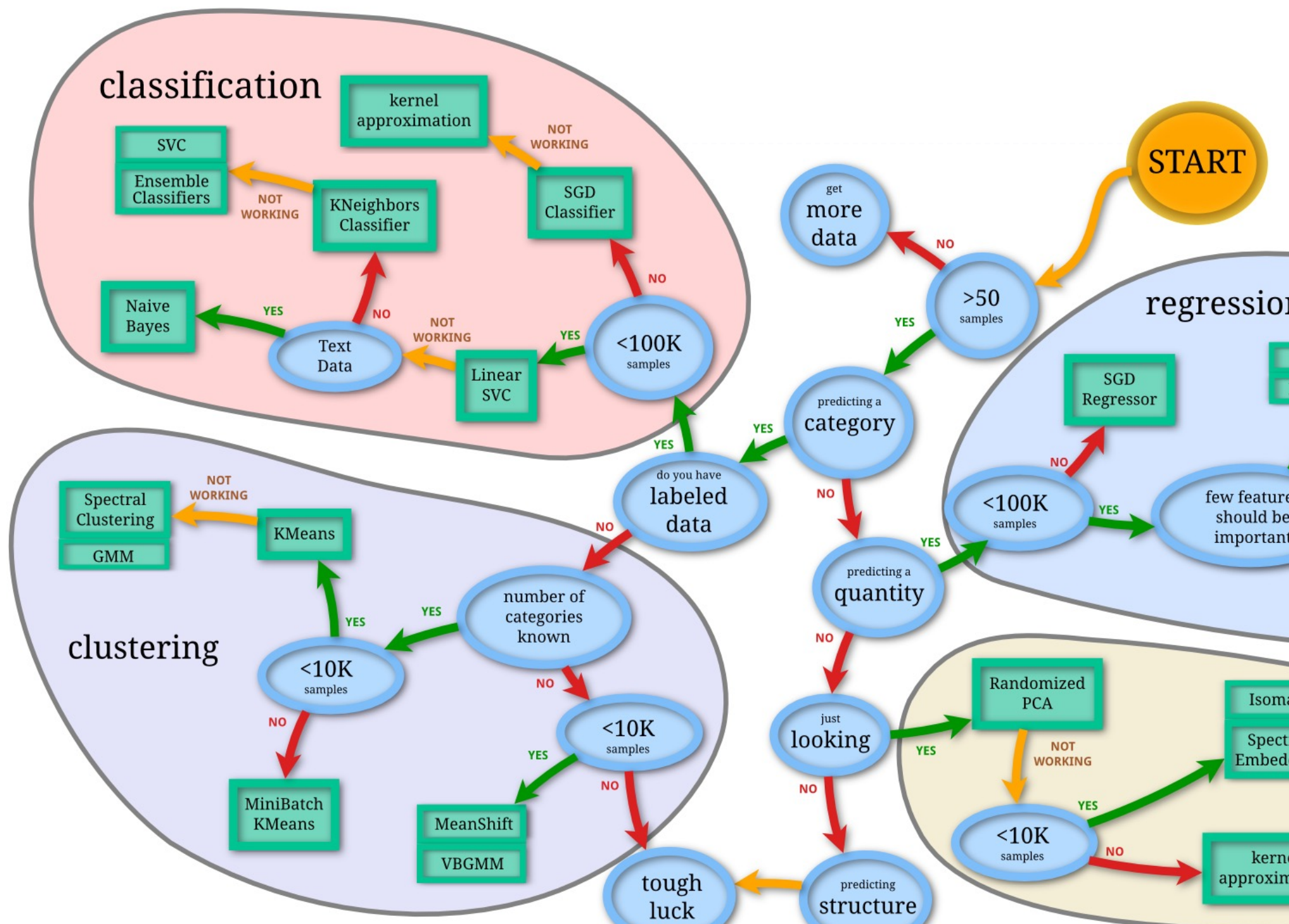
Python Machine Learning Package: Scikit Learn

- ▶ Most popular machine learning package for Python.
- ▶ It has a lot of algorithms built-in, such as:
 - ▶ Linear Regression
 - ▶ Logistic Regression
 - ▶ K Nearest Neighbors
 - ▶ Decision Tree
 - ▶ Random Forest
 - ▶ Support Vector Machine
 - ▶ K-Means Clustering
 - ▶ Principal Component Analysis
 - ▶ Naive Bayes
- ▶ Scikit Learn installation from terminal or Anaconda prompt, type: **conda install scikit-learn**, or **pip install scikit-learn**.



Machine Learning Map





Open your notebook, let's create our first
Machine Learning project