

# THE MACHINE LEARNING

## JOURNEY

### CHAPTER - 01



Harvard  
Business  
Review

Subscribe

Sign In

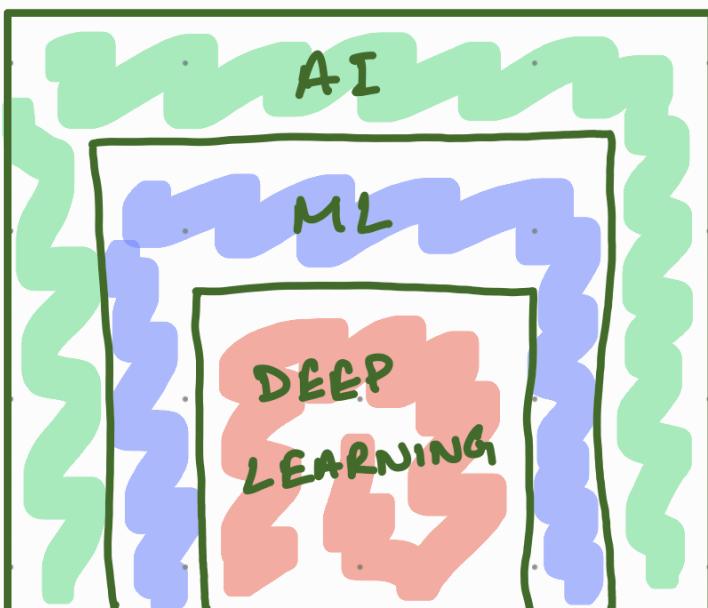


## Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)

@Shreyan



Machine Learning represents computer algorithms which can observe and analyse data on their own.

So speaking from an overview -

1. Collect data related to a particular area.
2. Develop an algorithm by which we can represent the relationship between various elements in the data.

Data is generally split into 2 parts:

- ↳ train
- ↳ test.

**train:** Train data is used to provide training to the model.

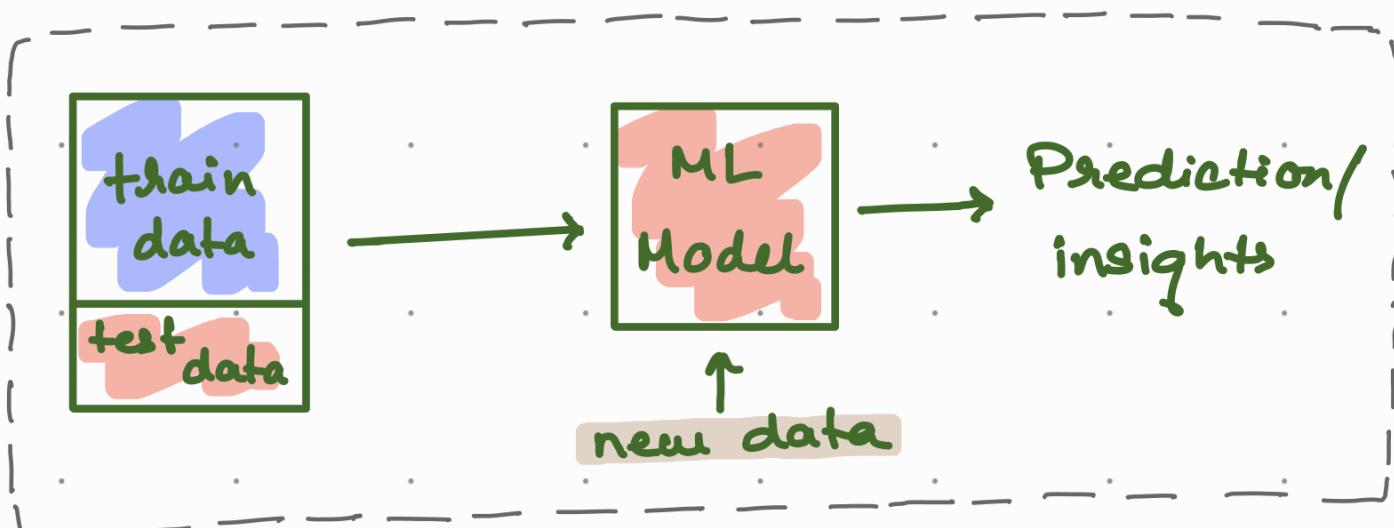
**test:** After training, the model will be able to predict the future behaviour with new data.

This means, we can test the model with the test data to check if it is working properly.

(@Shreyan)

So, after the model is developed, it will be able to work with new data that is not part of our already existing data.

Hence, the machine has learned to solve the problem.



Examples for application:

1. Predicting traffic conditions.
2. Product recommendations.
3. Voice recognition system.
4. Stock market predictions.
5. Spam email filtering.

# SUPERVISED Vs. UNSUPERVISED LEARNING

## (INTRO)

**Supervised model :**  $\phi$  takes data with column names and target value to be predicted.

$\phi$  data is generally divided into train and test data.

$\phi$  they understand the training data and check their learning accuracy using test data.

**eg -** from a dataset related to experience of employee and salary drawn.

So, for a new employee, depending on experience we can predict his salary.

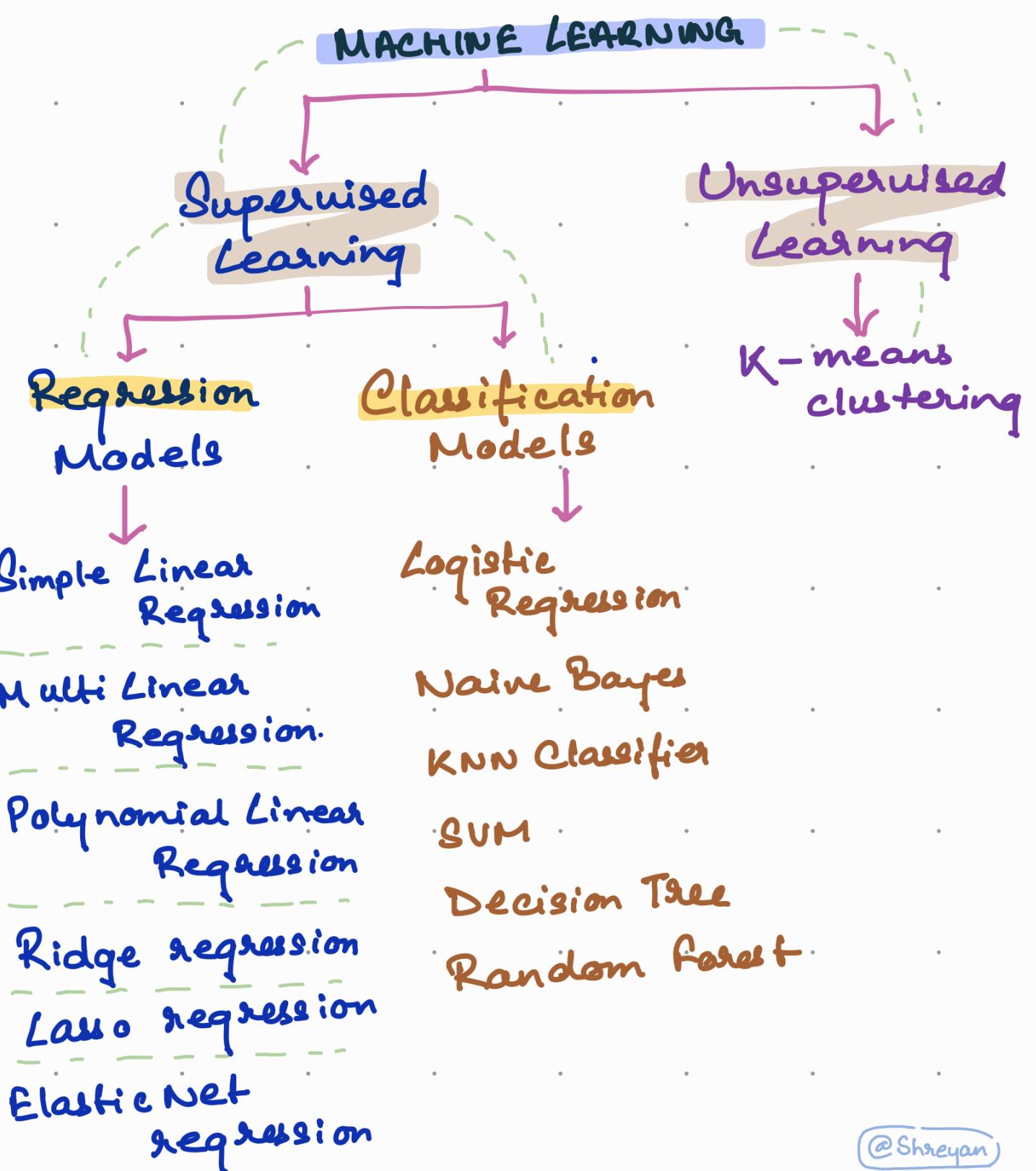
@Shreyan

**Unsupervised Model :**  $\phi$  they do not have data with labels and target columns.

$\phi$  data is understood by extracting features and patterns on their own.

$\phi$  does not make predictions but they provide some information about the given data.

**eg -** for items data in supermarket, we can classify the items as vegetables, oil, soaps depending on max similarity.



@Shreyan



SHREYAN BASU RAY  
OPENSOURCE MAINTAINER @SAGE.AI  
SUPPORT ME - [github.com/sponsors/Shreyan1](https://github.com/sponsors/Shreyan1)