**Tak 2 : My notes**

1.1

Machine learning is a subset of AI, Deep Learning is a subset of Machine Learning.

Ai: is a branch of Computer Science that is concerned with building smart & intelligent Machines.

Machine Learning: is a technique to implement AI that can learn from DATA by themselves without being explicitly programmed.

Deep Learning: is a subfield of ML that uses Artificial Neural Networks to learn from the data.

1.2

Machine Learning is a technique to implement AI that can learn from the data by themselves without being explicitly programmed.

Types of Machine Learning:

* Supervised Learning: the ML algorithm learns from **Labelled** Data
* Unsupervised Learning: the ML algorithm learns from **Unlabeled** Data
* Reinforcement Learning: is an area of ML concerned with how intelligent agents take actions in an environment to maximize its rewards
  1. Environment (chess board)
  2. Agent (computer)
  3. Action
  4. Reward (each step that come closer to win)

1.3

Types of Supervised Learning (2 main types)

* Classification: is about predicting a class or discrete values (Male or Female; True or False)
  + Algorithm:
    1. Decision Tree Classification
    2. Random Forest Classification
    3. K-nearest Neighbor
* Regression: is about predicting a quantity or continuous values (Salary; Age; price)
  + Algorithm:
    1. Logistic Regression
    2. Polynomial Regression
    3. Support Vector Machines

1.4

Types of Unsupervised Learning

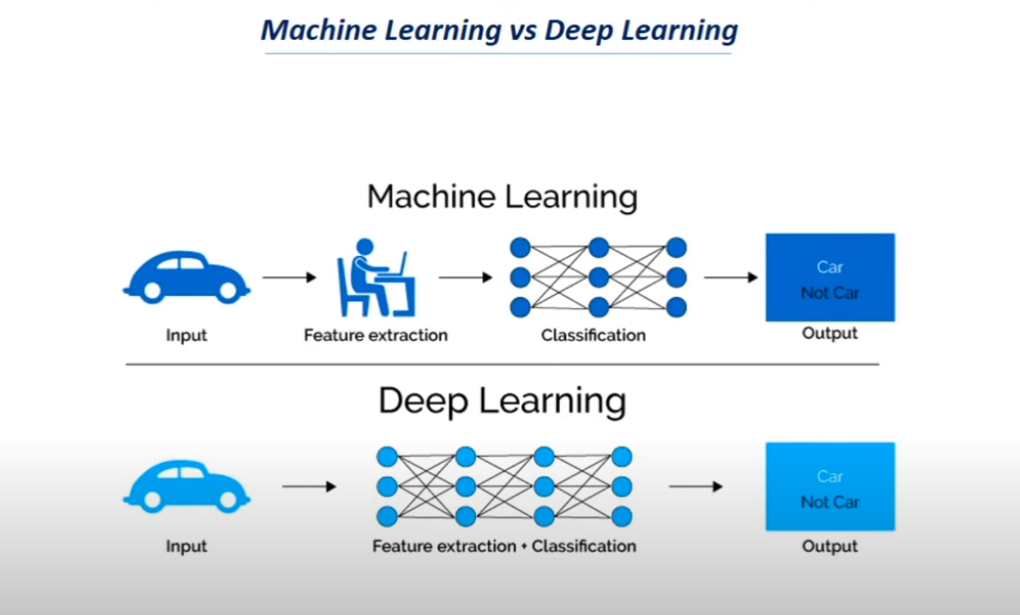
* Clustering: is an unsupervised task which involves grouping the similar data points
* Association: is an unsupervised task that is used to find important relationship between data points

Algorithm:

* K-means Clustering
* Hierarchical Clustering
* Principal Component Analysis (PCA)
* Apriori
* Eclat

1.5

Deep Learning is a subfield of Machine Learning that uses Artificial Neural Networks to learn from the data.

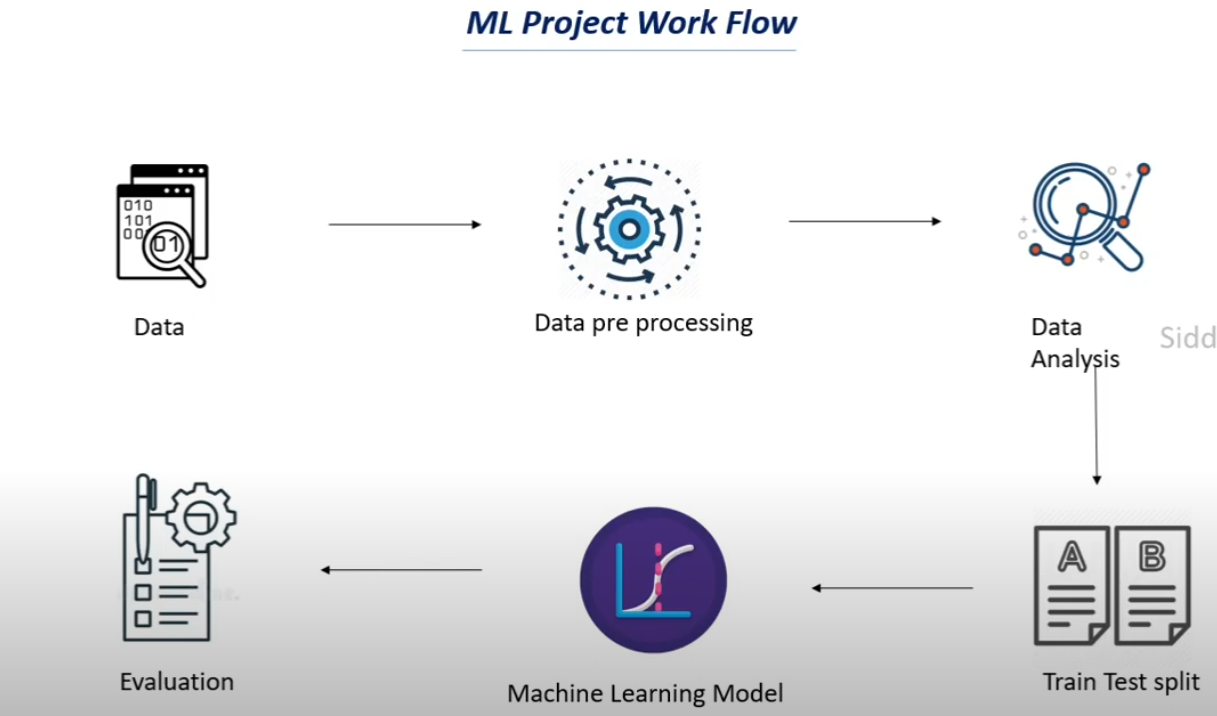


4.1

Where to collect Data?

* **Kaggle**
* **UCI Machine Learning Repository**
* **Google Dataset Search**

4.6



4.7

The mapping from textual data to real valued vectors is called feature extraction.

**Bag Of Words (BOW):** list od unique words in the text corpus

**Term Frequency-Inverse Document Frequency (TF-IDF):** To count the number of times each word appears in a document.

Term Frequency (TF) = (Number of times term t appears in a document) / (Number of terms in the document)

Inverse Document Frequency (IDF) = log-N/n), where N is the number of documents and n is the number of documents a term t has appeared in.

The IDF value of a rare word is high, whereas the IDF of a frequent word is low.

TF-IDF value of a term = TF \* IDF