# INTRODUCTION TO MACHINE LEARNING

## YASH RAJ OJHA
## ENG19CS0367

# What is Artificial Intelligence?

**AI**

**ML**

**DL**
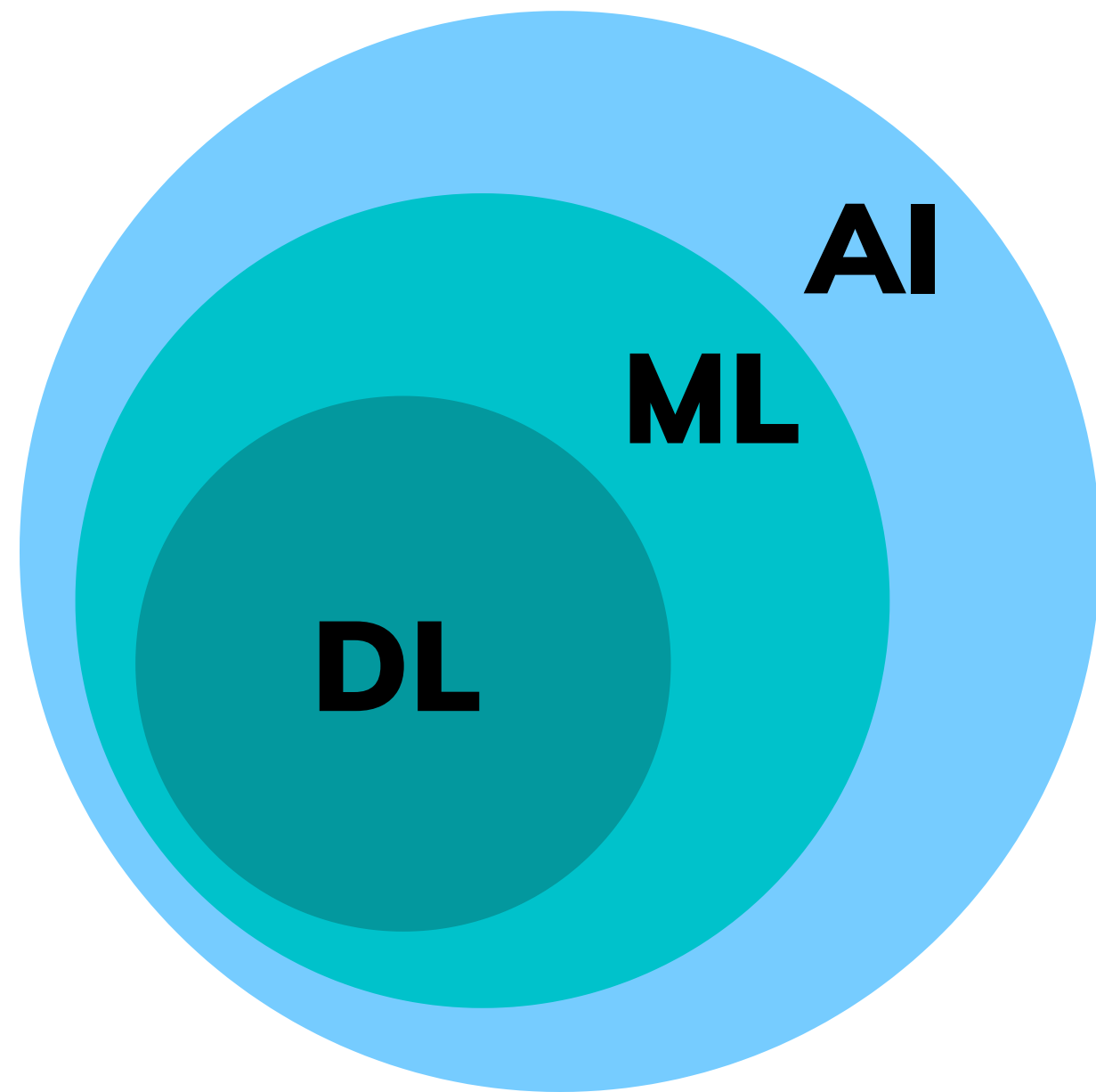
**Artificial intelligence (AI) is a branch of computer science concerned with smart machines that are capable of performing tasks that require human intelligence.**

- **Examples: Siri, Alexa and other smart assistants or Self-driving cars.**

# So what is Machine Learning?

**Get data**

↓

**Select ML Algorithm**

↓

**Give I/O data to Algo**

↓

**Model generation**

↓

**Predict output**

**Machine learning is a branch of AI that focuses on the use of data and algorithms to imitate the way that humans learn.**

**In normal program, we give the input and program to obtain a result,**

**In ML we generate a model with the help of an algorithm that accepts some inputs and outputs.**

→
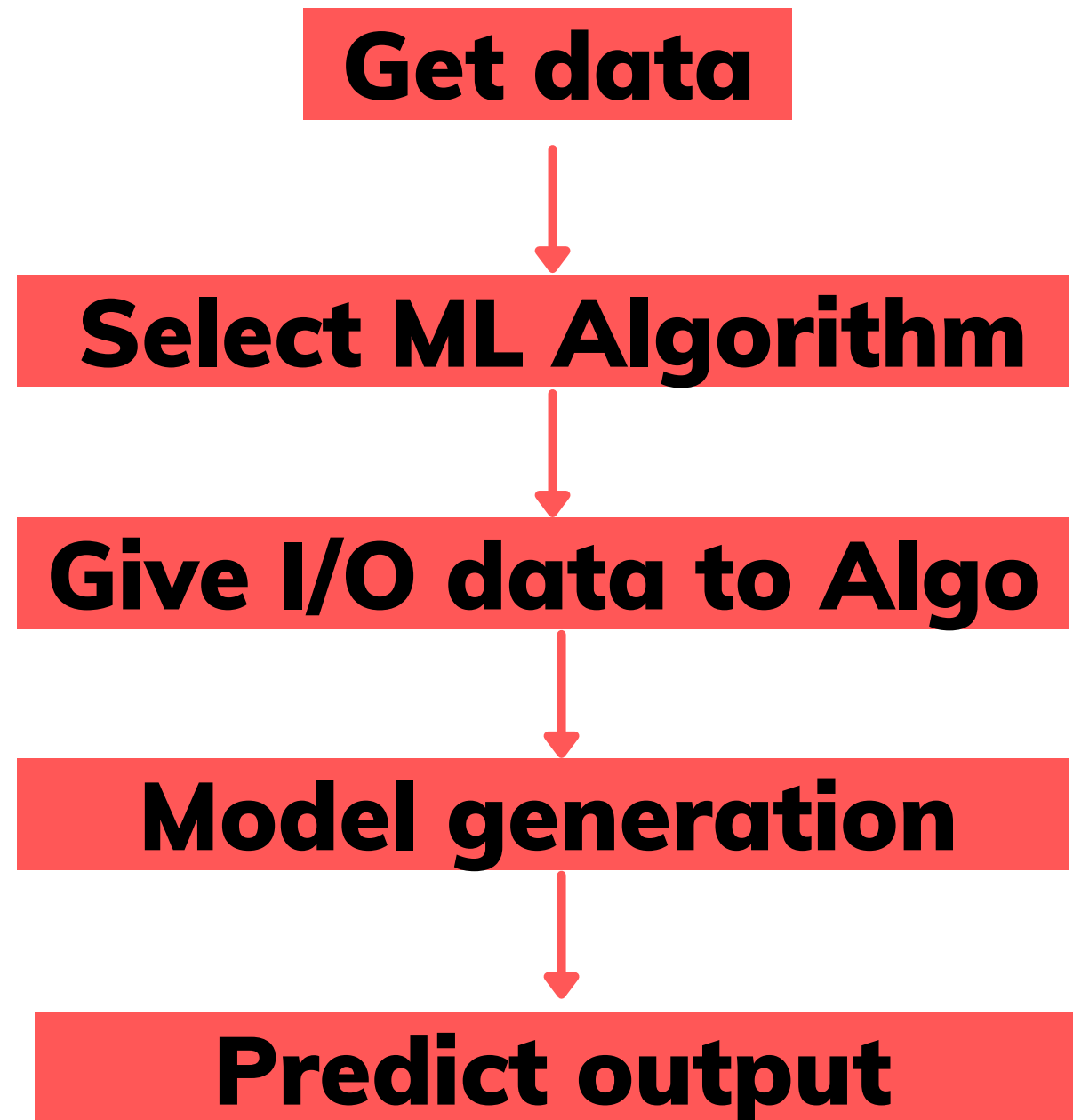
# Terminology In Detail

**As said before, the process of learning begins with analysing data, the algorithm understands the pattern in data and generates a model, which predicts output when we feed it with input.**

**Let's assume a function F(X) = Y**

**Here X is the <u>input</u>, it is called Features**
**Y is the <u>output</u> and it is called Labels**
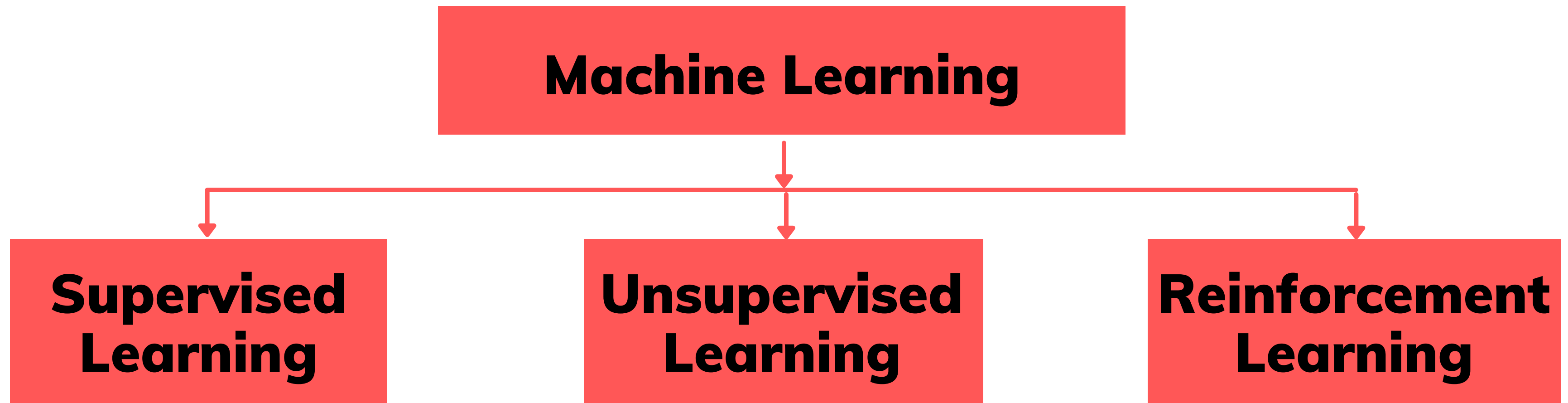**and <u>function F</u> is called Model/Hypothesis.**

→

# Example

Let's take an example of the fruit apple, now we know that the final product is the fruit which is apple so, in our case **apple** will be the **label**, it's **features** will be **diameter of the fruit**, the **size of fruit** and the **color of the fruit**.

Sometimes these are also called **attributes**. And the **number of features** are called **dimensions**. **Labels** are the final **output**. **Model** is like **real world process**.

→

# Types of Learning

**Machine Learning means machine learns something, and learning is classified into three types.**

```
                    ┌─────────────────────┐
                    │  Machine Learning   │
                    └─────────────────────┘
                               │
            ┌──────────────────┼──────────────────┐
            ▼                  ▼                  ▼
  ┌──────────────┐   ┌──────────────┐   ┌──────────────┐
  │  Supervised  │   │ Unsupervised │   │Reinforcement │
  │   Learning   │   │   Learning   │   │   Learning   │
  └──────────────┘   └──────────────┘   └──────────────┘
```

# Types of Learning

**1. Supervised Learning**: Supervised learning is when the model is getting trained on a labelled dataset. Labelled dataset is one which have both input and output parameters. In simpler words it means learning from someone. We have features and labels.

- **Example**: We teach a kid that dogs have 4 legs and a tail, they are more fierce than cats. Dogs bark while cats meow.

**2 .Unsupervised Learning**: When learning contains only some features without any description or labels, the algorithm will learn to distinguish on its own, it learns to identify a pattern on its own is called unsupervised learning.

→

# Types of Learning

- **Example**: Someone can group shapes like a square and triangle, based on the number of sides (features) without knowing what it is called.

**3. Reinforcement Learning**: The training of machine learning models to make a sequence of decisions based on output. It is reward based learning, we make decisions based on reward.

**Example**: You instruct the dog to get the ball and after he gets the ball he gets rewarded with a biscuit or else he gets beaten

→

# Types of Learning Simplified

**Supervised Learning**

> Labelled data
> Direct feedback
> Predict outcome

**Unsupervised Learning**

> No labelled data
> No direct feedback
> Find hidden pattern in data

**Reinforcement Learning**

> Decision process
> Reward system
> Learn from series of actions

# Types of Supervised Learning

**Supervised learning can be separated into two types of problems, Classification and Regression :**

**1. <u>Classification</u>: In this method the algorithm recognizes specific entities and attempts to draw some conclusions on how those entities should be labeled or defined. It involves grouping the data.**

**Example : If we get more than 40 marks then it is pass, if its less than 40 then its fail, a person is male or female, does a person have diabetes or not.**

→

# Types of Supervised Learning

**2. <u>Regression</u>:** In regression, a single output value is produced using training data. The data we provide is continuous.

**Example**: If we tell 1pen is 10rs, 2pen is 20rs, 3pen is 30rs so 8 pen is ? The model will predict 80rs. This method is used predict the price of some product, like predicting value of houses.

→

# Types of Unsupervised Learning

**Unsupervised learning can be separated into two types of problems, Clustering and Association:**

**1. Clustering: "Clustering" is the process of grouping similar entities together. The goal of this unsupervised machine learning technique is to find similarities in the data point and group similar data points together. output labels are not known beforehand.**

**Example : Groups of photos with similar cars, types of news.**

# Types of Unsupervised Learning

2. **<u>Association</u>**:  This technique checks for the dependency of one data item on another data item and maps accordingly so that it can be more profitable. It uses the concept of association.

**Example** : Bread in a supermarket is generally kept with jam, butter or a phone cover is generally purchased with a phone

→

# Data Mining and Data Warehouse

**<u>Data Mining</u>: Data mining refers to extracting or mining knowledge from large amountsof data. The data is processed using ML algorithms to make decisions. We can use scikit learn.**

**<u>Data Warehouse</u>: A data warehouse is a large collection of business data used to help an organization make decisions. We get the data from different sources, each data will be of different format, we extract the data and transform it. After cleaning the data we load it into database. It contains in relevant data.**

# Data Warehousing & Data Mining

## Data Warehousing & Data Mining Relationship

The diagram below outlines the relationship between Data Warehousing and Data Mining:
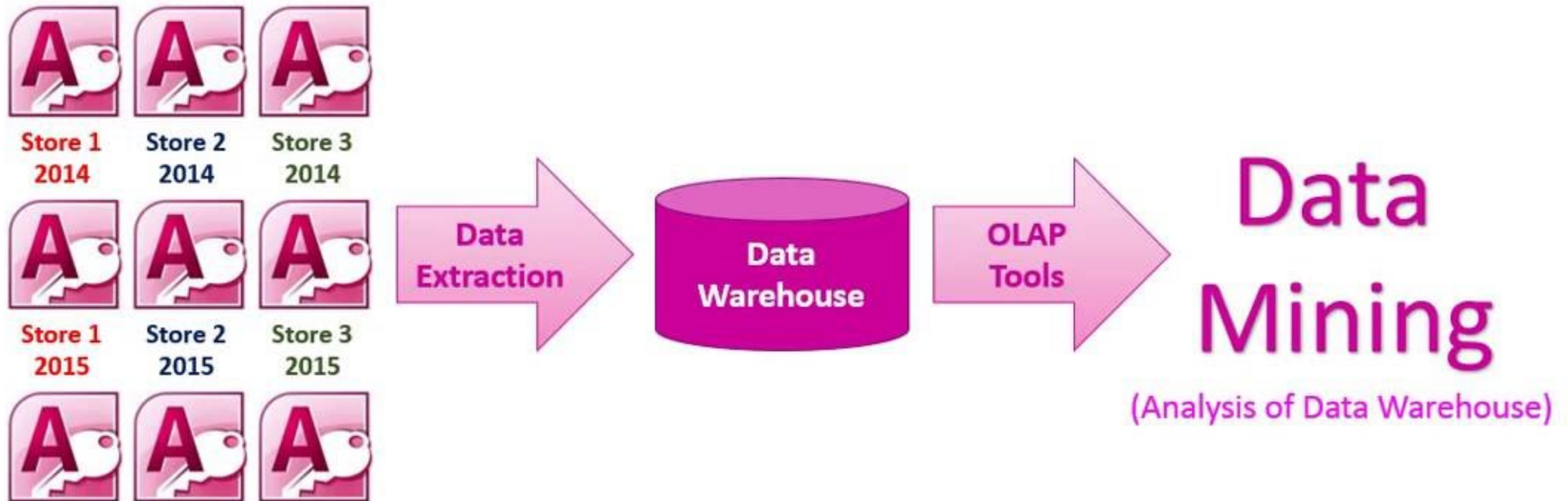
# Three steps involved in ML

**1. ML programming requires 3 lines of code. We <span style="color:red">decide on a learning algorithm</span>.**

**2. We implement the algorithm inside a class, by writing a <span style="color:red">fit method</span> which we call. This fit method requires input and output which generates a model when called.**

**3. When we call <span style="color:red">predict method</span> it predicts the output when we give it input.**

→

# Common Machine Learning Algorithms

1. **<u>Linear Regression</u>**: It is used to estimate real values (cost of houses, number of calls, total sales etc.) based on a continuous variable(s). Here, we establish a relationship between independent and dependent variables by fitting the best line. The equation is in form of **y=mx+c**.

**Example**: If we give features and labels as 1 = 2, 2 = 4, 3=6, 4=8, then 5=?, our model will predict 10.

**Code in GitHub and Google Colab.**

$\rightarrow$

# Common Machine Learning Algorithms

**2. Logistic Regression: It is a classification algorithm, It is used to estimate discrete values ( Binary values like 0/1, yes/no, true/false ) based on a given set of the independent variables (s). In simple words, it predicts the probability of occurrence of an event.**

**Example: If marks > 40% it is pass, or fail, we have only two answers to a question, yes or no. To predict whether an email is spam (1) or (0).**

**Code in GitHub and Google Colab.**

→

# Common Machine Learning Algorithms

3. **Decision Trees**:  In this technique, we split the population or sample into two or more homogeneous sets (or sub-populations) based on a differentiator question in input variables. The decision rules are generally in form of if-then-else statements. The deeper the tree, the more complex the rules and fitter the model.

**Example**: Invest in stock 1 or 2, Is the movie good or not?

Code in GitHub and Google Colab.

→

# Common Machine Learning Algorithms

**4. Random Forest**:  A random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. The (random forest) algorithm establishes the outcome based on the predictions of the decision trees.

**Example**: Should we invest in stock 1 or 2? Which stock is good for the long term, which stock will yield more benefits?

Code in GitHub and Google Colab.

# Common Machine Learning Algorithms

**5. Support Vector Machine: It is majorly used for classification algorithms. Classify depending upon features and put them based on one side of the plane or the other side of the plane.
We perform classification by finding the hyper-plane that differentiates the two classes very well.**

**Example:  We see a strange cat that also has some features of dogs, we use SVM to differentiate.**

**Code in GitHub and Google Colab.**

# Common Machine Learning Algorithms

6. **K-Nearest-Neighbors**: It is used for the classification problems, The algorithm assumes that similar things exist in close proximity. K-NN algorithm stores all the available data and classifies a new data point based on the similarity.

**Example**:  If a new person comes, take his coordinates and find the closest group and put him in that group.

Code in GitHub and Google Colab.

→

# Applications of Machine Learning

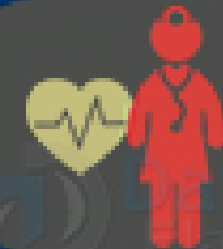## Machine Learning Applications in Healthcare

- **Machine Learning Examples in Healthcare for Drug Discovery**: Manufacturing or discovering a new drug is a lengthy process as thousands of compounds need to be tested. Machine learning can speed up one or more of these steps in this lengthy multi-step process.

- **Personalized Treatment/Medication**:  Machine learning could play a vital role in finding what kind of genetic makers and genes respond to a particular treatment or medication
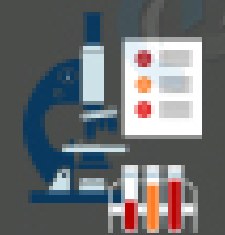
→

# Machine Learning Applications in Healthcare



- Diseases Identification & Diagnosis
- Drug Discovery & Manufacturing
- Medical Imaging
- Personalized Medicine/ Treatment
- Smart Health Records
- Diseases Prediction

# Applications of Machine Learning

## Machine Learning Applications in Finance

-**Machine Learning Examples in Finance for Fraud Detection**: With increasing online shopping, fraudulent transactions are also on the rise. To combat this, companies are using machine learning algorithms to identify and block fraudsters

- **Machine Learning Applications in Retail**: Retailers are implementing big data technologies to build big data solutions. They need a solution that can analyse the data in real-time and provide valuable insights that can translate into tangible outcomes like repeat purchasing.