Course Name: Machine Learning for Beginners

Basics of Machine Learning

- Supervised Machine Learning
- Unsupervised Machine Learning
- Reinforcement Learning

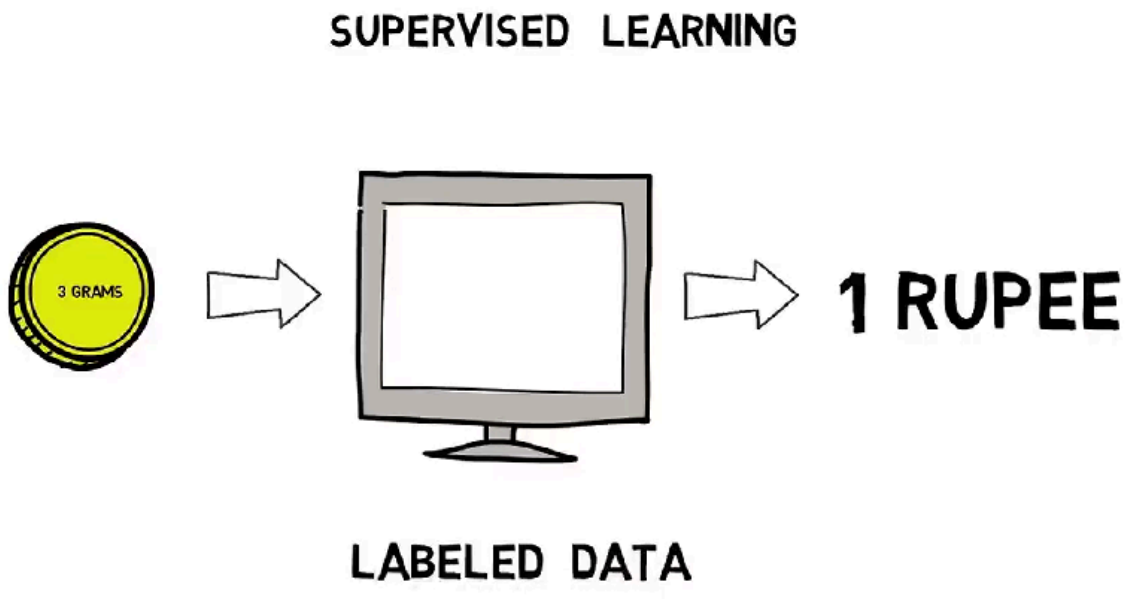
Supervised Machine Learning:

- Suppose you have 3 coins. These are

Here

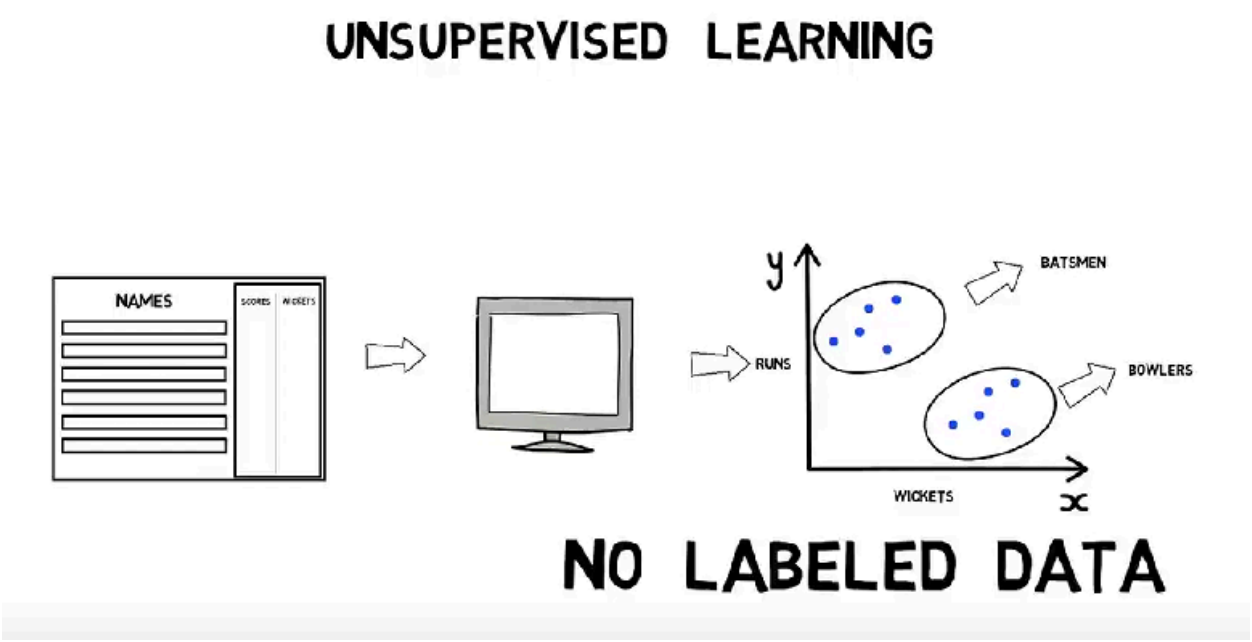
- Weight = Feature
- Currency = Label

It will learn if coins = 3 grm, then it will say 1 rupee



Unsupervised Machine Learning

- Unlabeled data means unsupervised machine learning.
- Here's an example of cricketer runs and wickets.



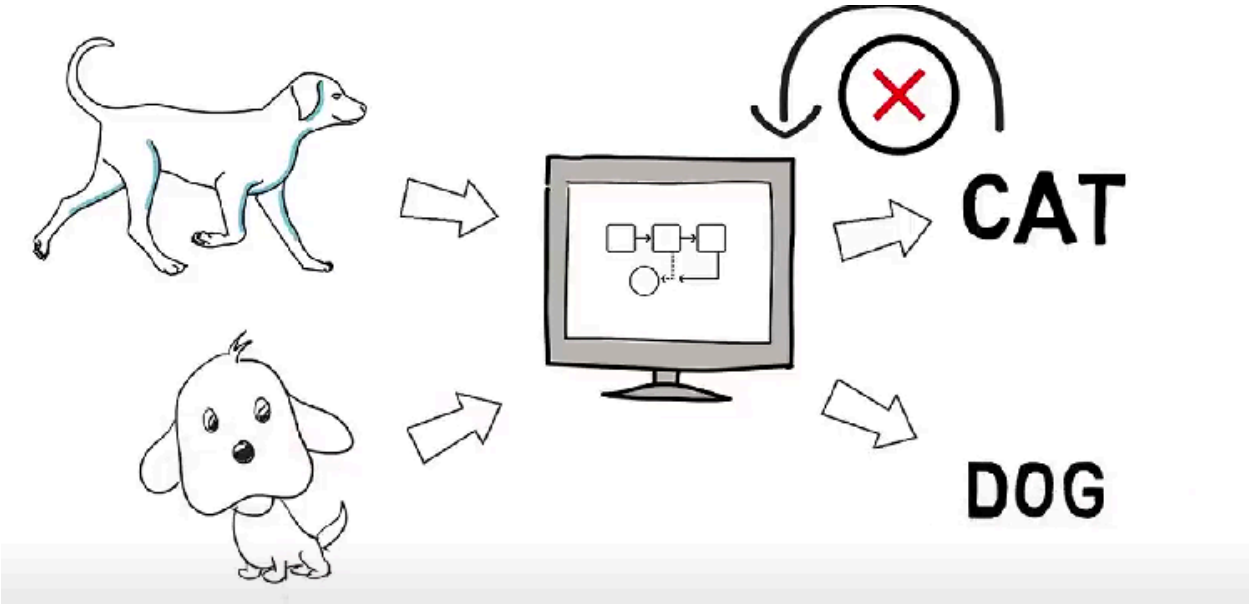
There was 2 cluster

- first one is took more runs but less wickets
- The second one took more wickets and fewer runs.

Here, no label is provided with the data, so this is called unsupervised learning.

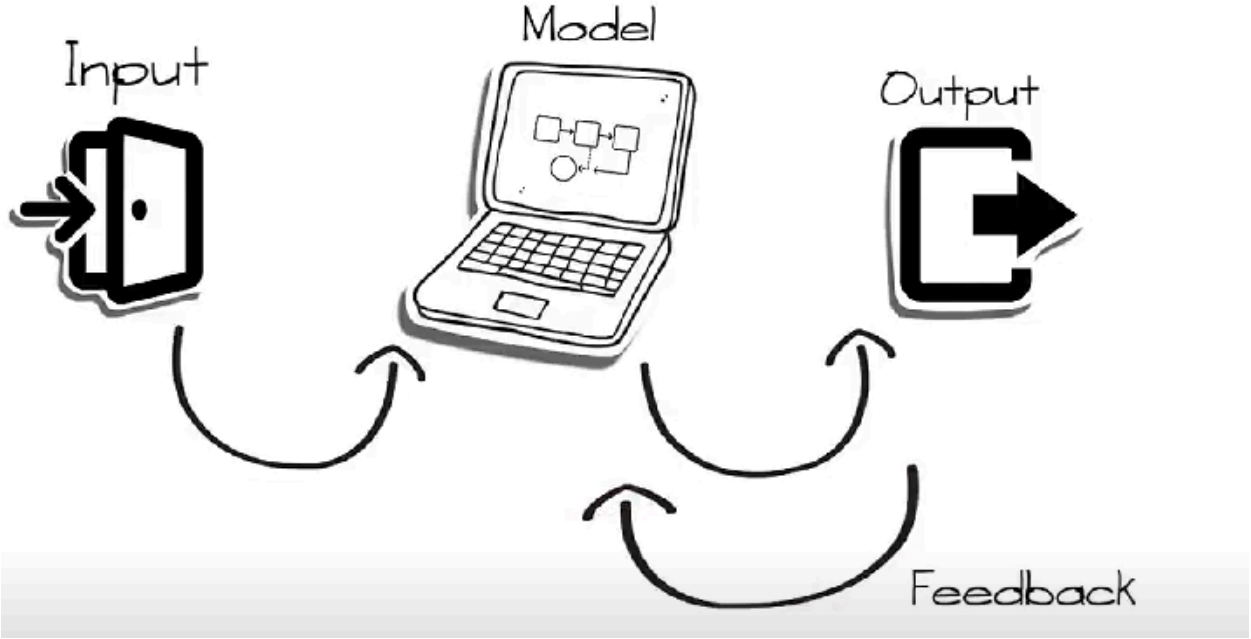
Reinforcement Learning

- Reward-based learning: let's say I was given an image that is a dog, then the machine says it is a cat, then I can give negative feedback. The machine gets this feedback and learns from this feedback. After I am given another image, if it is a dog image, then it predicts or identifies it correctly. This is called reinforcement learning.



Machine Learning Model

- We give an input to the machine learning model, and then it gives an output according to the algorithm we applied in this model.
- We take the output if it is correct; otherwise, we give feedback to learn the model.



How Is Machine Learning Possible in This Era?

- We used google to search for specific info, the search words are data.
- We buy so many things from e-commerce; it stores our interested products and also data.

So, these data are used to mainly train ML models to learn human interest or other things.

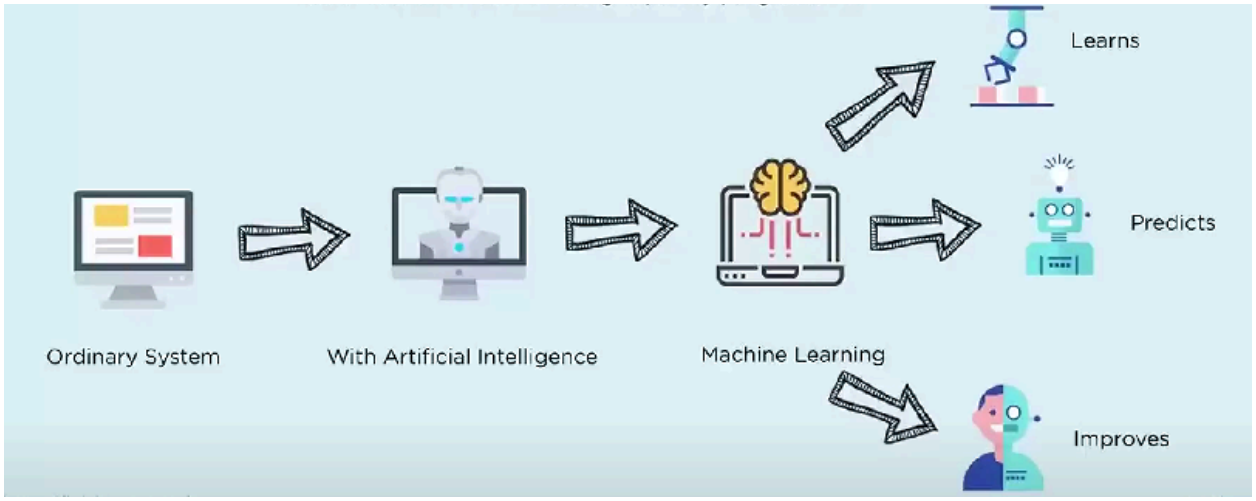
Applications of Machine Learning:

- Healthcare
- Sentiment Analysis
- Fraud detection
- Customer Choice
- Pricing of any product

Why Machine Learning

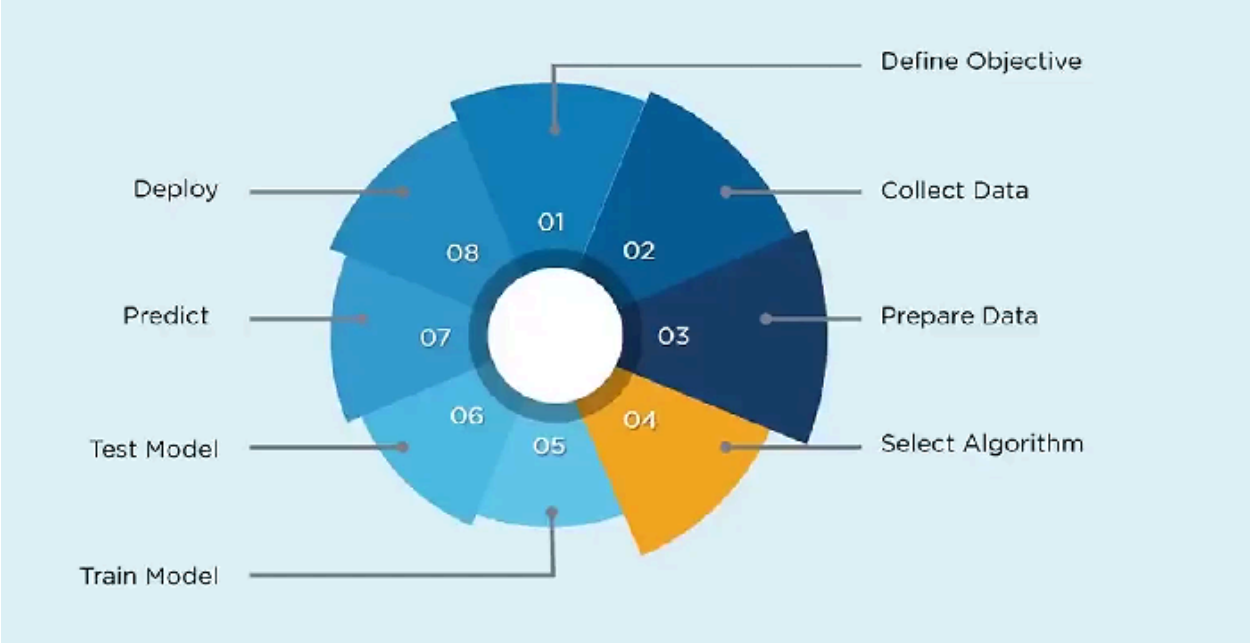
- Machine Can drive
- Detect Eye Disease
- Unlock phone used face

Machine learning is the science of making computers learn and act like humans by feeding them data and information without being explicitly programmed.



What is machine learning?

- Define Objective : What is it we want to detect?
- Collect Data : Collect the relevant data
- Prepare Data : Prepared for Data to predict
- Select Algorithm : Select algorithms for predict, like SVM (support vector machine)
- Train Model : Train the model using the datasets
- Test Model : Check the model worked correctly.
- Predict : Use sample to predict sample.
- Deploy : After lots of testing, deploy the model.

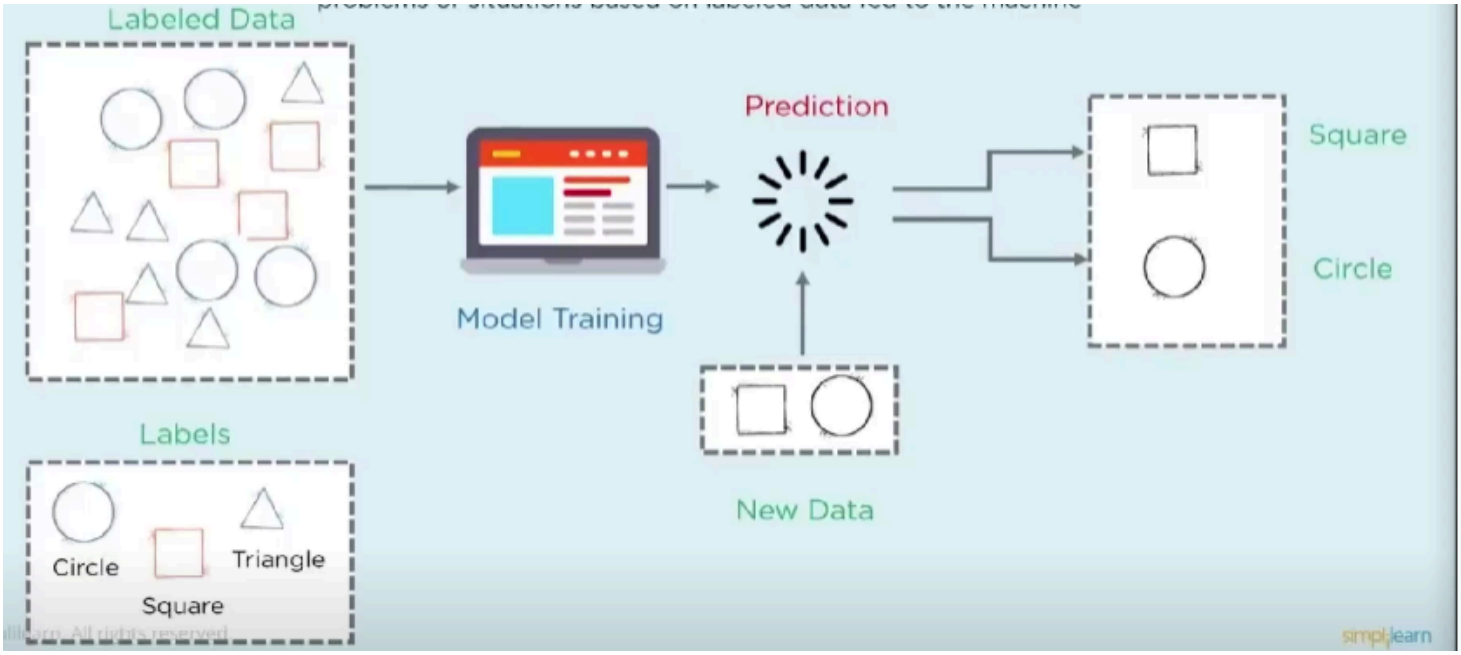


What does machine learning do?

- For category -> Classification
- For quantity -> Regression
- For Anomaly -> anomaly Detection
- For Untrai/Structure -> Clustering

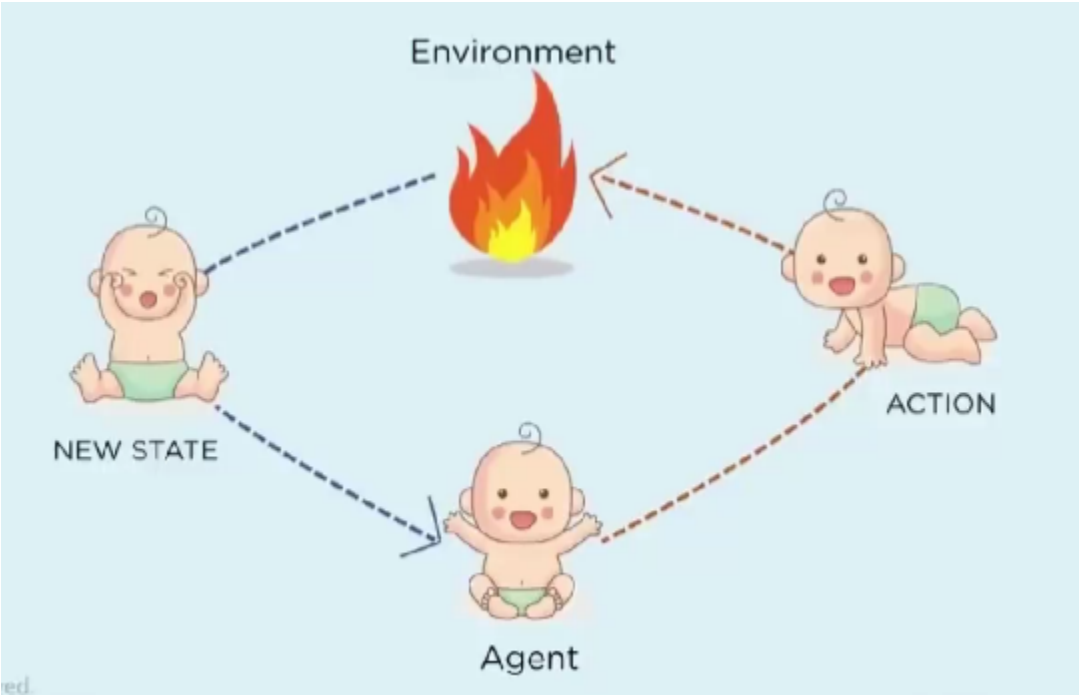
Types of Machine Learning

- Supervised learning is a method to enable machines to classify/predict objects, problems, or situations based on labeled data fed to the machine.



Reinforcement Learning

- Reinforcement learning is an important types of ML where an agent learns how to behave in an environment by performing actions and seeing the result.



Supervised Vs Unsupervised

Machine Learning Algorithms

- Linear Regression
- Decision Tree
- Support Vector Machine

Linear Regression:

- Linear regression is perhaps one of the most well-known and well-understood algorithms in statistics and machine learning.
- Linear regression is a linear model.
- Linear relation between the input variable (x) and the single output variable (y)
    - Y = mx + c

Suppose we are predicting distance traveled (y) from speed (x). Our linear regression model representation for this problem would be

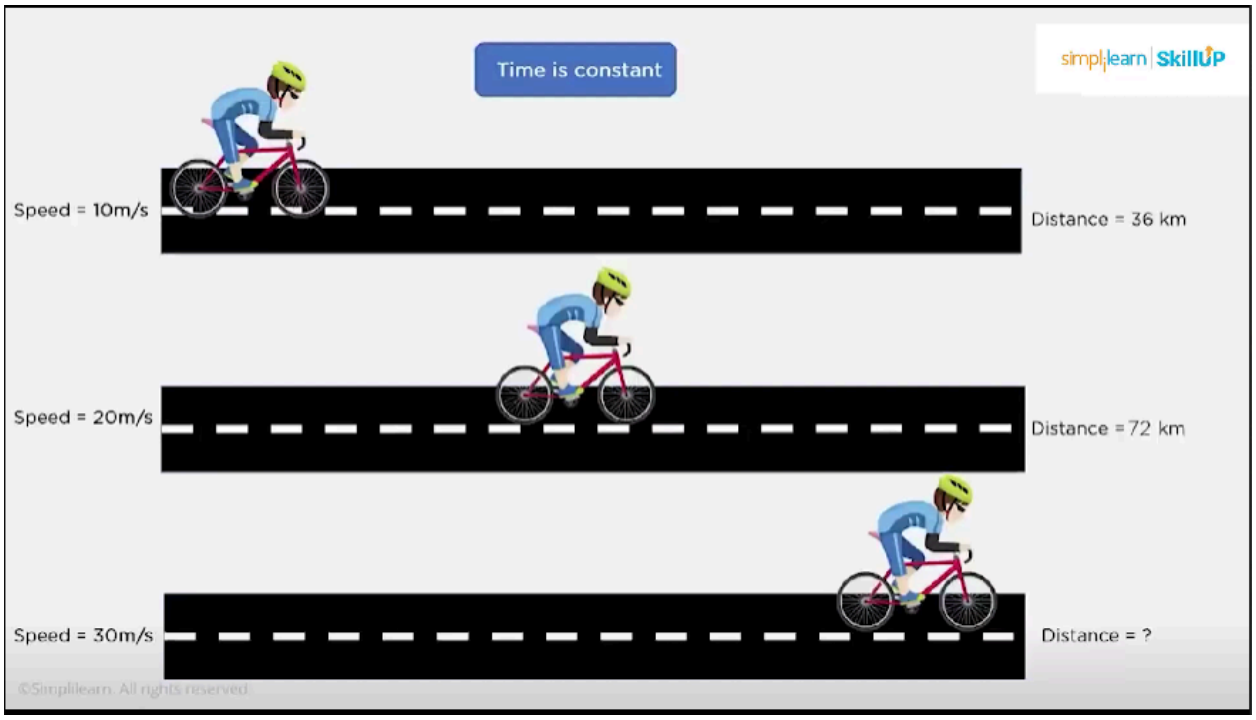
Y = m*x + c

Where,

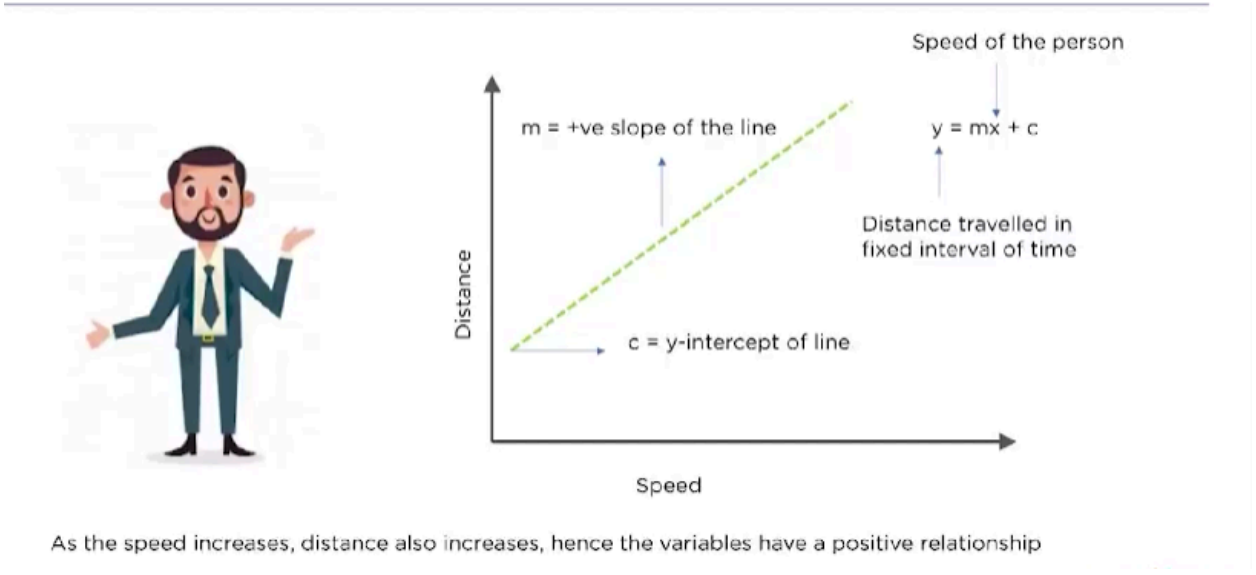
Distance = m * speed + c

C = coefficient

M = y-intercept.

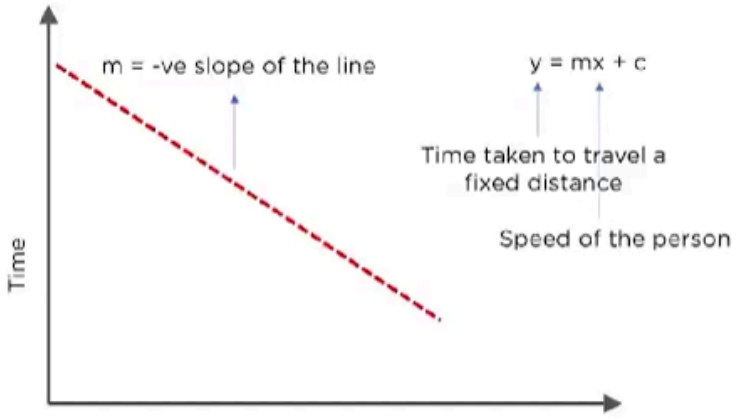


Linear Regression: Distance against Speed



Speed against Distance

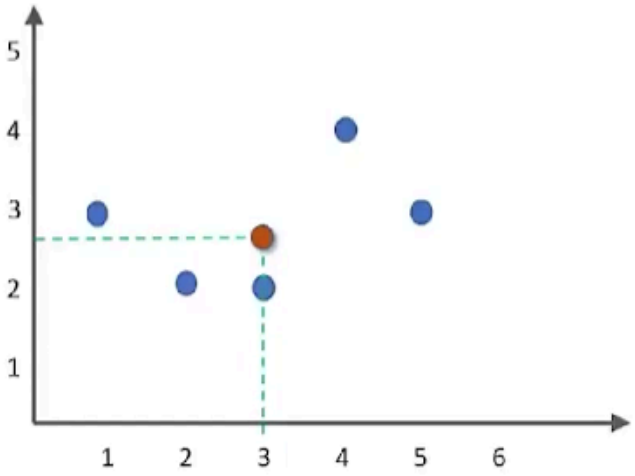
If distance is assumed to be constant, let's see the relationship between speed and time



Let's plot these points!!

x	y
1	3
2	2
3	2
4	4
5	3

Mean(xi) = 3 Mean(yi) = 2.8



$y = mx + c$

$$m = \frac{\sum (x - xi)(y - yi)}{\sum (x - xi)^2}$$

x	y	x- xi	y - yi	(x- xi) ²	(x- xi) (y - yi)
1	3	-2	0.2	4	-0.4
2	2	-1	-0.8	1	0.8
3	2	0	-0.8	0	0
4	4	1	1.2	1	1.2
5	3	2	0.2	4	0.4
				Total = 10	Total = 2

So, we can calculate the value of c

$$y = mx + c$$

$$m = \frac{\sum (x - xi)(y - yi)}{\sum (x - xi)^2} = 2/10 = 0.2$$

$y = 0.2\,x + c$ Mean values = (3, 2.8)

$2.8 = 0.2 * 3 + c$

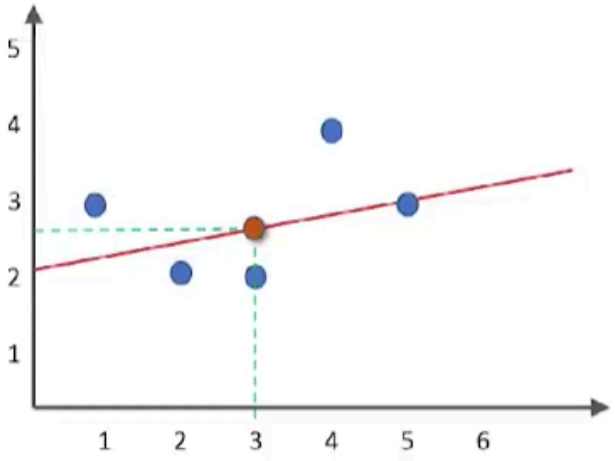
$2.8 = 0.6 + c$

$c = 2.8 - 0.6$

$c = 2.2$

Hence this is our regression line!

$y = (0.2 * x) + 2.2$



Now, let's predict the values of y using x = {1,2,3,4,5} and plot the points!

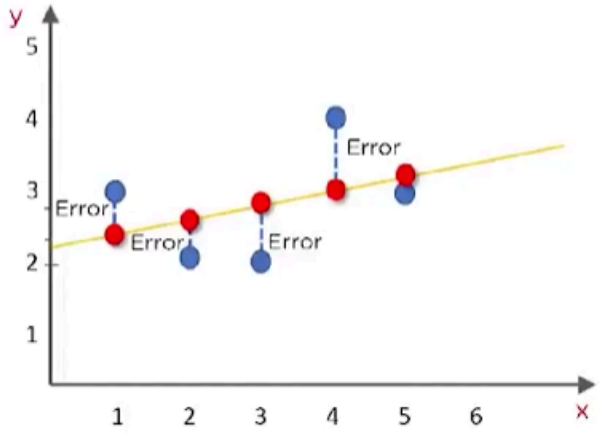
$y = (0.2 * x) + 2.2$

$y_p = (0.2 * 1) + 2.2 = 2.4$
$y_p = (0.2 * 2) + 2.2 = 2.6$
$y_p = (0.2 * 3) + 2.2 = 2.8$
$y_p = (0.2 * 4) + 2.2 = 3.0$
$y_p = (0.2 * 5) + 2.2 = 3.2$

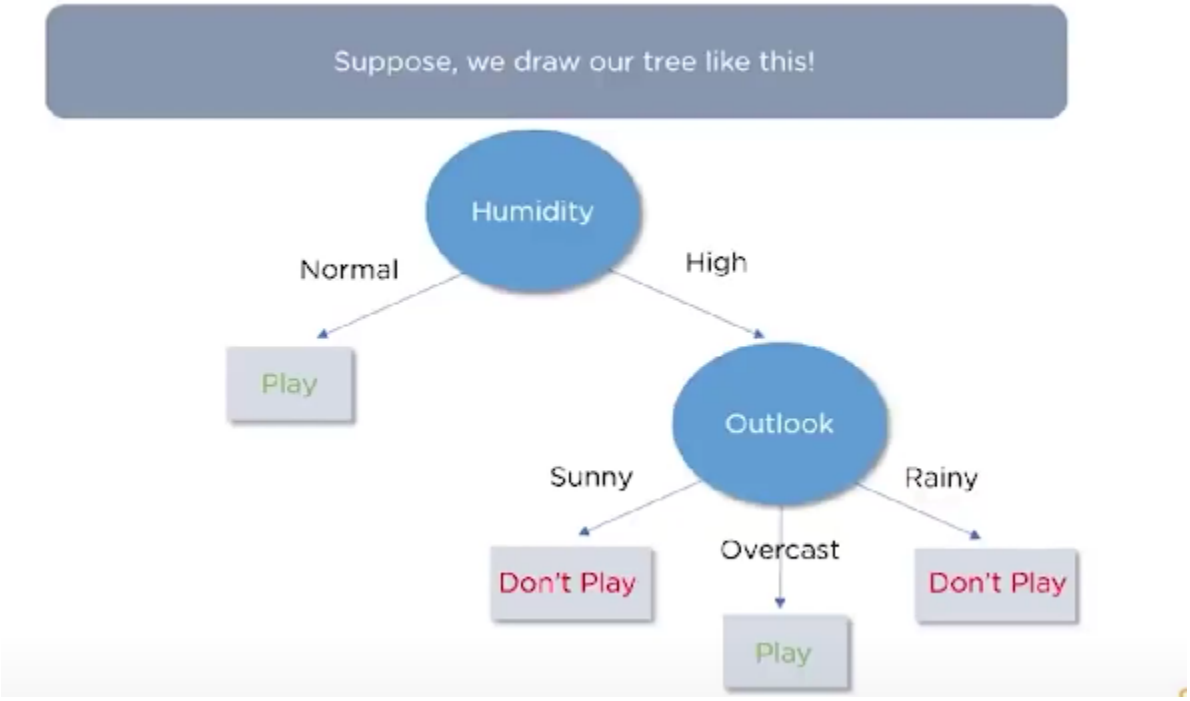
y_p = Predicted values of y

Plot the predicted values along with the actual values to see the difference

x	y	y_p
1	3	2.4
2	2	2.6
3	2	2.8
4	4	3
5	3	3.2



Minimizing the distance : There are lots of ways to minimize the distance between the line and the data points, like sum of squared errors, sum of absolute errors, Root Mean Square error, etc.



Decision Trees

- It is a flowchart structure used for both classification and regression tasks in machine learning. It visually represents a decision-making process, breaking down complex data into simpler, manageable parts. The tree starts at a root node and branches out based on decision rules, leading to final outcomes or classifications.

Entropy

- Is the measure of randomness or 'impurity' in the datasets

Let's look at entropy:

$$Entropy = I(p,n) = -\frac{p}{p+n} \times log_2(\frac{p}{p+n}) - \frac{n}{p+n} \times log_2(\frac{n}{p+n})$$

Let's look at entropy!

a) Entropy of target class of the dataset (whole entropy):

Entropy (Play golf)

= E(5,9)

= I(5/14, 9/14)

= I(0.36, 0.64)

= -(0.36 $\log_2$ 0.36) - (0.64 $\log_2$ 0.64)

= 0.94

Play Golf	
Yes	No
9	5

Total = 14

Let's look at entropy!

Entropy (Play golf, Outlook)

= P(sunny) * E (3,2) + P(Overcast) * E(4,0) + P(rainy) * E(2,3)

= 5/14 * I(3,2) + 4/14 * I(4,0) + 5/14 * I(2,3)

= 0.693

Similarly, we can calculate the entropy of other predictors like Temperature, Humidity, Windy!

		Play Golf		
Predictors		Yes	No	Total
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

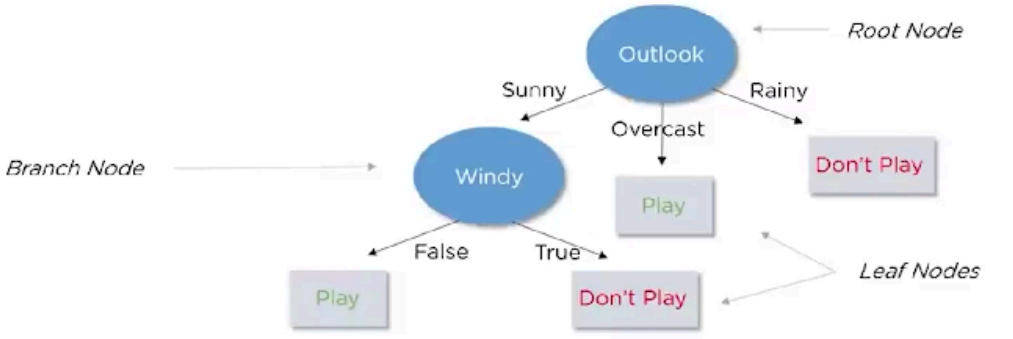
Gain(Outlook) = Entropy(PlayGolf) − Entropy(PlayGolf,Outlook)

= 0.940−0.693

=0.247

The information gain of the other three attributes can be calculated in the same way:

Gain(Temp)	= Entropy(PlayGolf)−Entropy(PlayGolf,Temp)	= 0.029
Gain(Humidity)	= Entropy(PlayGolf)−Entropy(PlayGolf,Humidity)	= 0.152
Gain(Windy)	= Entropy(PlayGolf)−Entropy(PlayGolf,Windy)	= 0.048



We choose the attribute with largest information gain as the root node

Information Gain

- It is the measure of decrease in entropy after the dataset is split. Also known as entropy repudiations.

Support Vector Machine

- is a supervised machine learning algorithm used for both classification and regression tasks. It's particularly effective in high-dimensional spaces and aims to find the best hyperplane (a decision boundary) that separates data into different classes by maximizing the margin between the closest data points of different classes.

Datasets:

Let's have a look at our dataset:

Type	Flour	Milk	Sugar	Butter	Egg	Baking Powder	Vanilla	Salt
Muffin	55	28	3	7	5	2	0	0
Muffin	47	24	12	6	9	1	0	0
Muffin	47	23	18	6	4	1	0	0
Muffin	45	11	17	17	8	1	0	0
Muffin	50	25	12	6	5	2	1	0
Muffin	55	27	3	7	5	2	1	0
Muffin	54	27	7	5	5	2	0	0
Muffin	47	26	10	10	4	1	0	0
Muffin	50	17	17	8	6	1	0	0
Muffin	50	17	17	11	4	1	0	0
Cupcake	39	0	26	19	14	1	1	0
Cupcake	42	21	16	10	8	3	0	0
Cupcake	34	17	20	20	5	2	1	0
Cupcake	39	13	17	19	10	1	1	0
Cupcake	38	15	23	15	8	0	1	0
Cupcake	42	18	25	9	5	1	0	0
Cupcake	36	14	21	14	11	2	1	0
Cupcake	38	15	31	8	6	1	1	0
Cupcake	36	16	24	12	9	1	1	0
Cupcake	34	17	23	11	13	0	1	0

Implementation in Python:

```
# Packages for analysis# Packag
import pandas as pd
import numpy as np
from sklearn import svm

# Packages for visuals
import matplotlib.pyplot as plt
import seaborn as sns; sns.set(font_scale=1.2)

# Allows charts to appear in the notebook
%matplotlib inline

# Pickle package
import pickle
```

Application of Machine Learning:

- Virtual Personal Assistance
- Traffic Predictions
- Social Media Personalization
- Email Spam Filtering
- Online Fraud Detection
- Stock Price Prediction
- Diagnose Diseases
- Automatic Translation

What Does a Machine Learning Engineer Do?

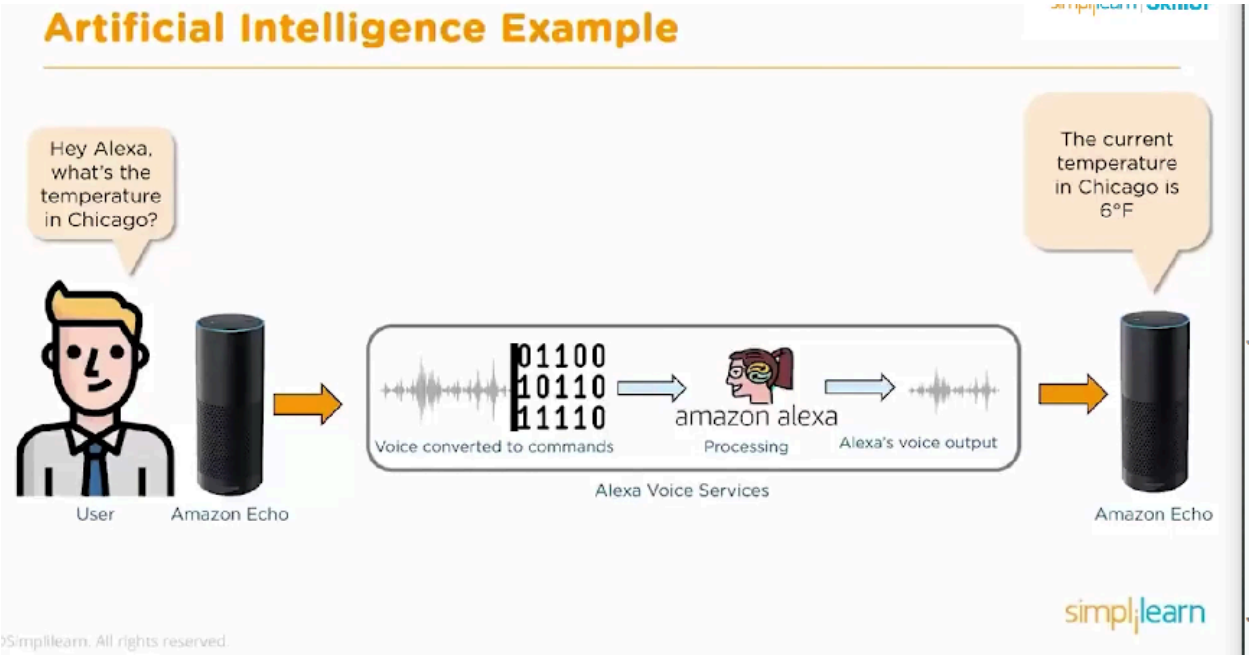
- Optimized these solutions for performance and scalability.
- Creating and maintaining machine learning solutions to solve business
- Solved business problems like reducing customer churn, running targeted marketing campaigns, and improving product experiences.

Learning Path:

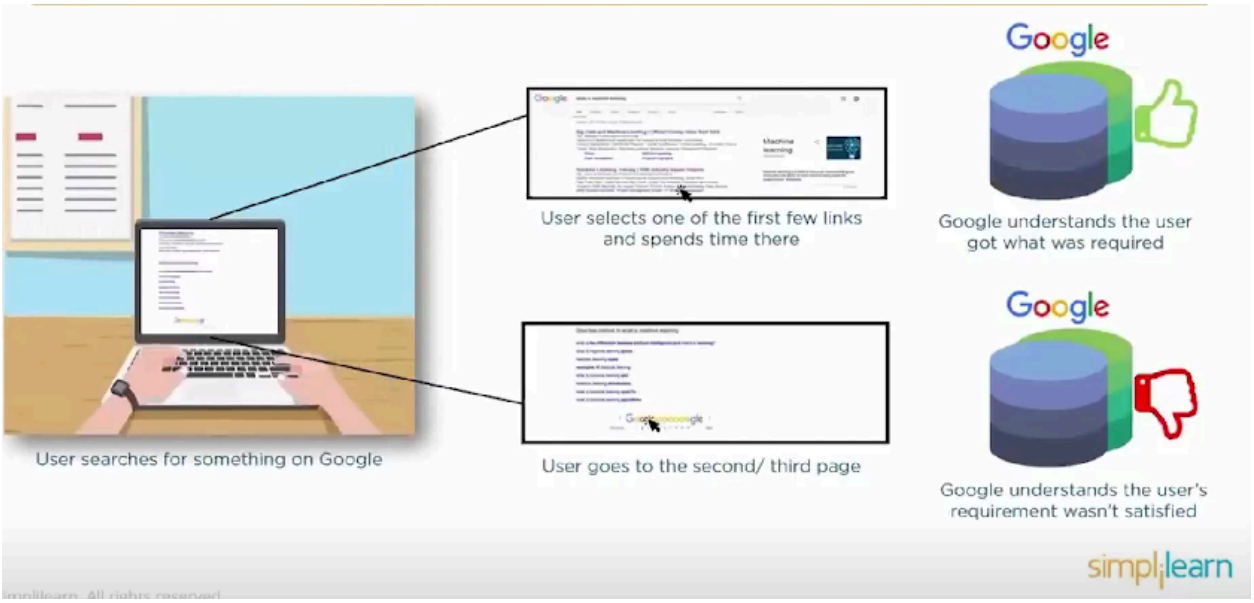
- Math : Probability , Calculus, Linear Algebra
- Programming Skills: Python, C, C++, C#.
- Data Engineering Skills: Preprocessing, Extract Transform Load, DBMS
- ML Algorithms : Supervised, Unsupervised
- Machine Learning Framework : TensorFlow, scikit-learn.

ML Vs DL Vs AI

Artificial Intelligence:



Machine Learning: Google is one of them.



Deep Learning: