Agenda :-

1. What is Data?
2. What is Information?
3. Types of Data.
4. What is ML?
5. Types of ML.
6. Common Apps of ML
7. Data platforms for finding datasets.

24 April 2023 19:45

1. What is Data?

85

85 % - Score of a student in an exam.

8. Odd no.
9. Temp
eg:- 1. Count of some object
10. runs of batsman 2. Score
3. age
4. %
5. marks
6. int
7. Special.

85 - abstract → Data is values

85% - conclusion.

↓

Information

↓

Data with some context.

85 runs

&

Values can either be a numerical or text value or image or audio

24 April 2023 19:52

Data - Values — [Numerical
Text

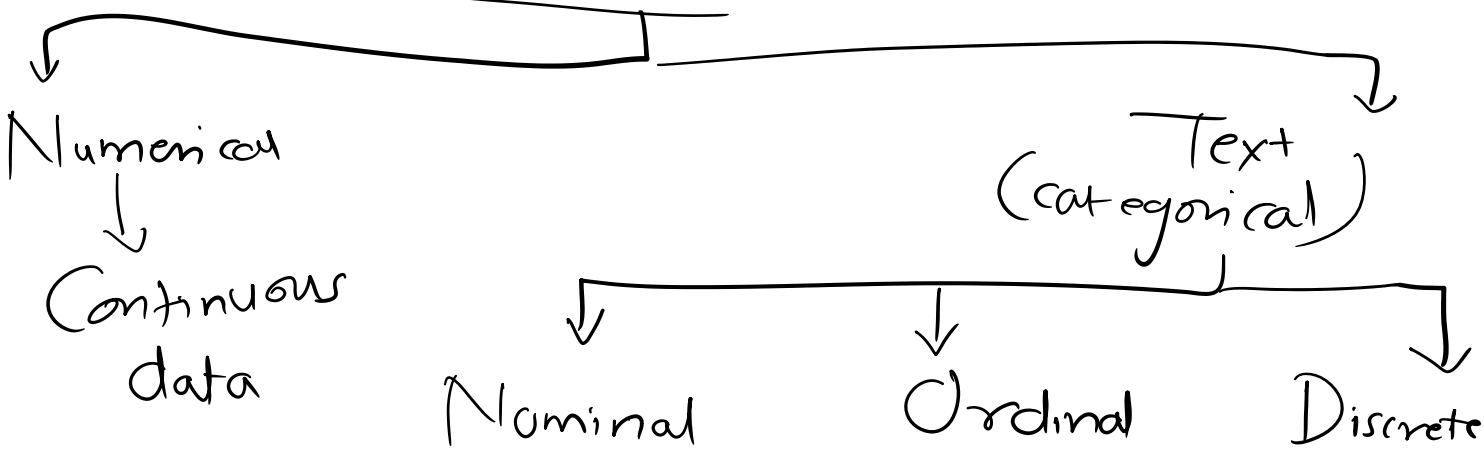
Paris
↓
Data

Numerical value - 85

Text value - Paris

place Paris, France
hollywood Paris Hilton
actor name

3. Type of Data :-



Continuous Data :- eg :- Height & Weight

↓

A data on which you can apply meaningful arithematic operations like avg, sum, std dev.....

	Height	Weight
1st	165.3cm	57.3kg
2nd	172.4cm	60.5kg

165.32 cm

165.325 cm

165 cm 166cm

Categorical Data :-

Nominal	Ordinal	Discrete
eg: ① State of a country ② Name of all employees.	eg:- ① Grades of a Student in an exam like A, B, C, etc. ② Reviews of a restaurant like Good, Bad, Avg, Poor.	eg :- ① Rating of a product like 1, 2, 3, 4, 5 etc ② NPS rating 1, 2, 3. ..... etc.

24 April 2023 20:08

Nominal :- Data has no specific order
(No-minal) eg :- Name of all emp
(No-order) Name of States of a country

X X X

Ordinal :- Data here has a specific order
(Order) eg :- Grades :- A, B, C, D, E
(Ord-inal) Rating :- Good, Bad, Poor.

Discrete :- A numerical data on which you cannot perform any mathematical operations like avg, sum, std_dev.....

eg:- 1 2 3 4 5

1.5 3.75 64

3 4

3.7 3 4

0, 40, 60, 80, 100

↓

Numerical

↓

Cannot apply any arithematic operations on this data.

100 − 100

80 − 100 − 80

60 − 80 − 60

40 − 60 − 40

0 − 40 = 0

Age :- 35 or 36 or 34

35.5436

Discrete :- ① No decimal values
② Whole Nos. 1, 2, 3,

24 April 2023 20:27

☆ :Note :- Most of times discrete data will be "id's" colums.

eg:- Emp.ID
Passenger ID
Student ID

1
2
.
.
.
99
100

avg of EmpID

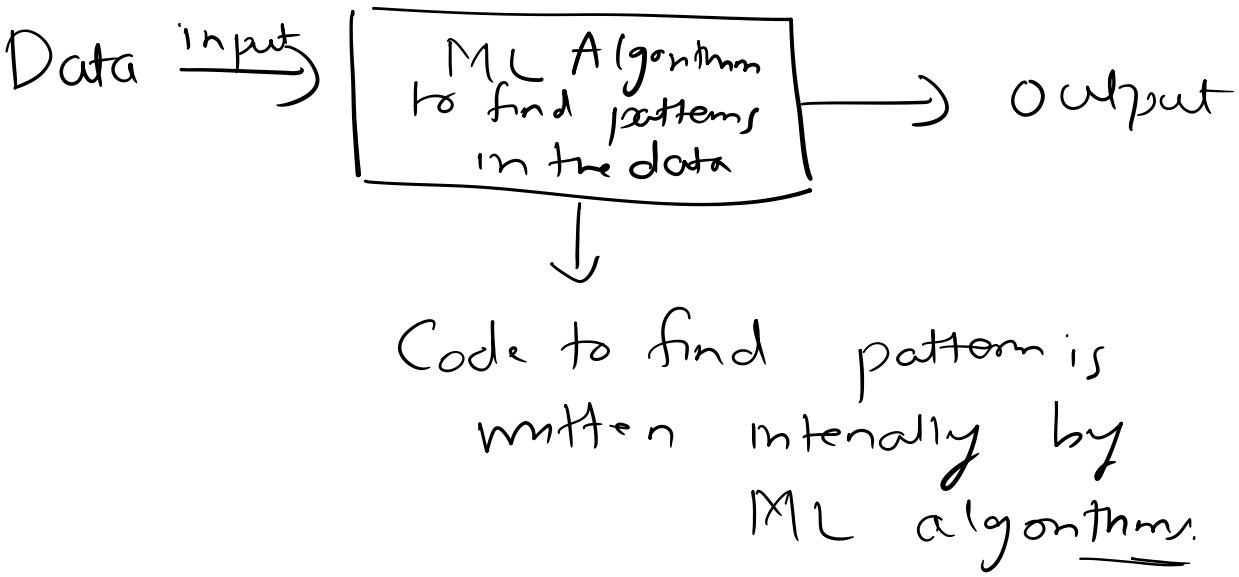
100 / 2 = 50

What is Machine Learning?

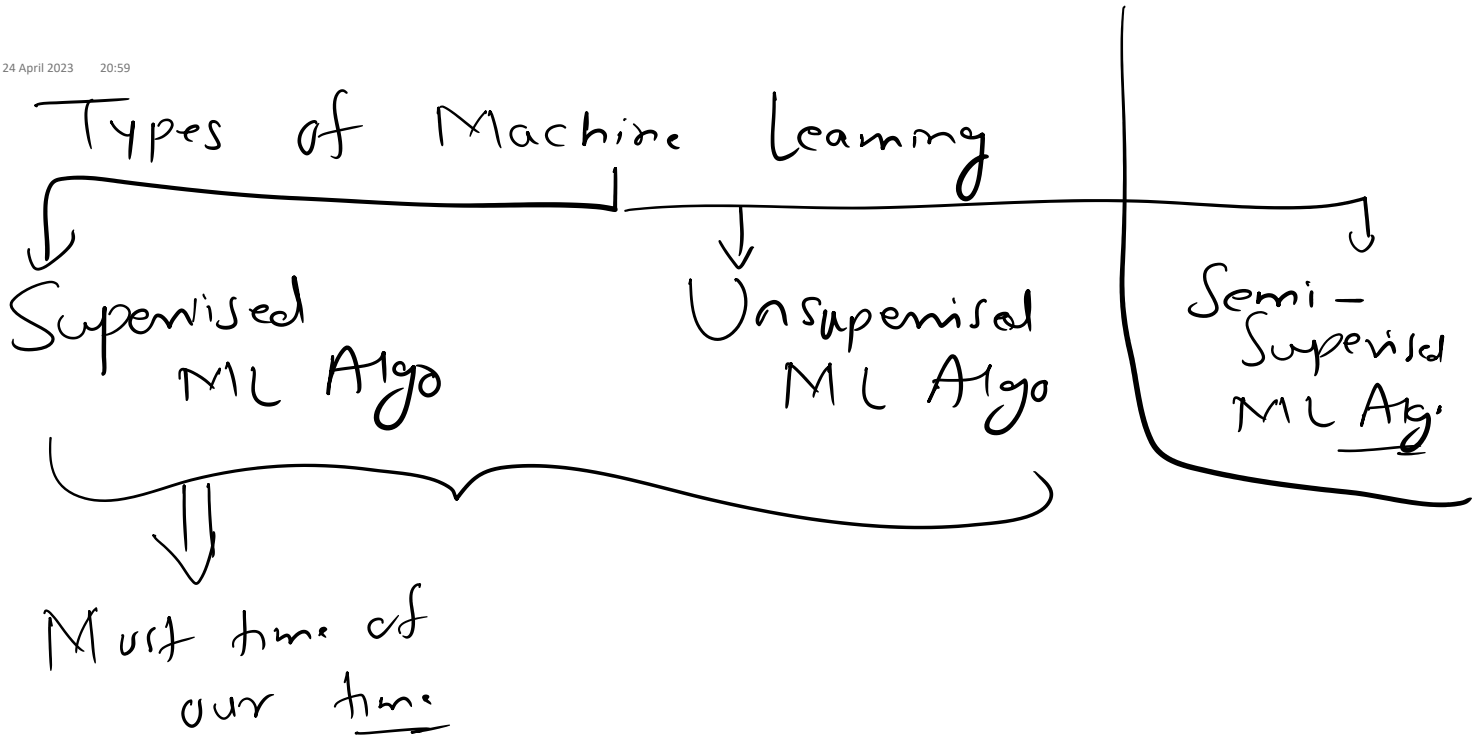
Before ML :-

Data —input→ Programmer will write code to find patterns from the data ⟶ Output

With Machine Learning :-



24 April 2023 20:59



24 April 2023 21:00

Variable / Features / Columns

In a dataset, there a 2 types of Variables

- Independent Variables
- Dependent Variable.

Yrs of Exp	Salary
10	2,50,000
5	1,50,000
⋮	⋮

Salary :-
Dependent Variable

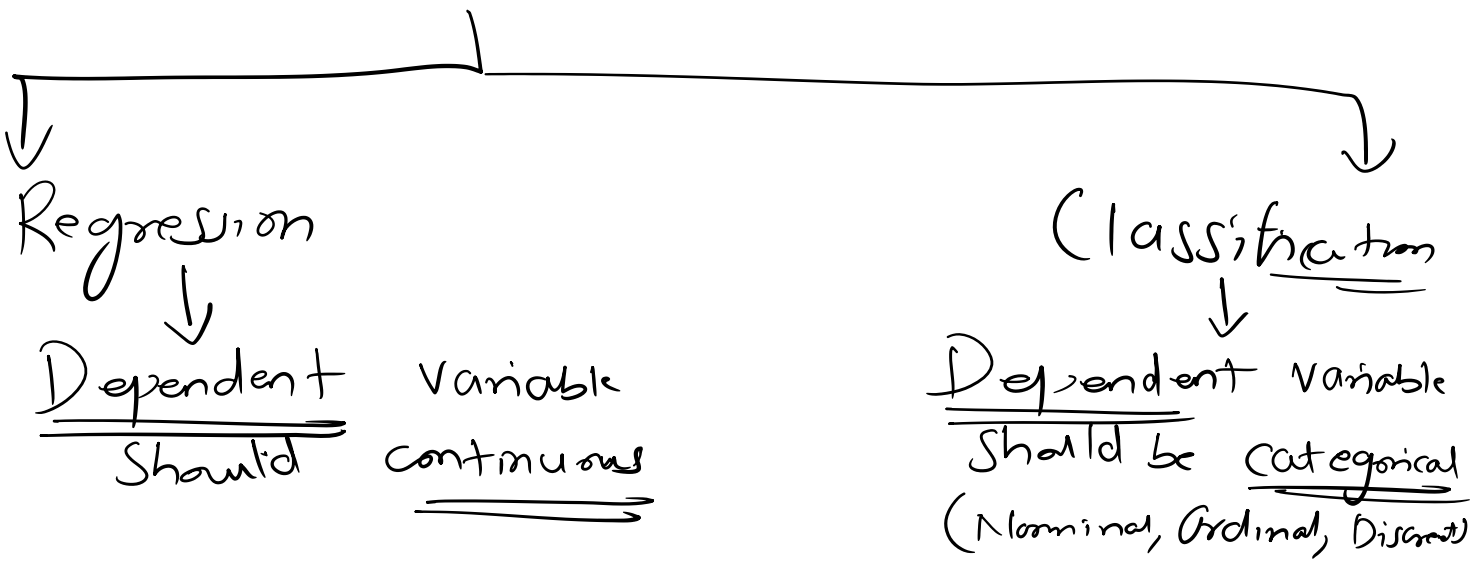
Yrs of Exp :-
Indepent Variable

ML process :-

① Dependent Variable (In a dataset there will be only 1 dependent variable & many Independent Variables)

② Always find the data types of columns

① Supervised :-



Regression :- Linear Reg, Decision Tree Regressor,
Random Forest Regressor, Knn Regressor,
SVM Regressor, etc.

Classification :- Log. Reg, Decision Tree Classifier,
Random forest Classifier, Knn Classifier
SVM Classifier, NB Classifier

24 April 2023 21:18

Common Applications of ML Algo :-

1. ChatGPT
2. Amazon Product Recommendation
3. Alexa, Siri
4. Self-Driving Cars.
5. Netflix Movie Recommendation

(2) UnSupervised ML Algorithm :-

- KMeans Clustering,
- DBSCAN,
- Hierarchical Clustering

☆. No Dependent Variable

(3) Semi Supervised ML Algorithm :-

Reasearch

(1) Online Jupyter Notebook :- Google Colab.