WsCube Tech — ML Part-1

10:00 Am
saturday
19-04-2025

Youtube → Playlist
Mail → spam
Amazon Shopping → Bought together

} Because of machine learning

What is machine learning?

Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and informati in the form of observations and real-world interactions.

Ex:

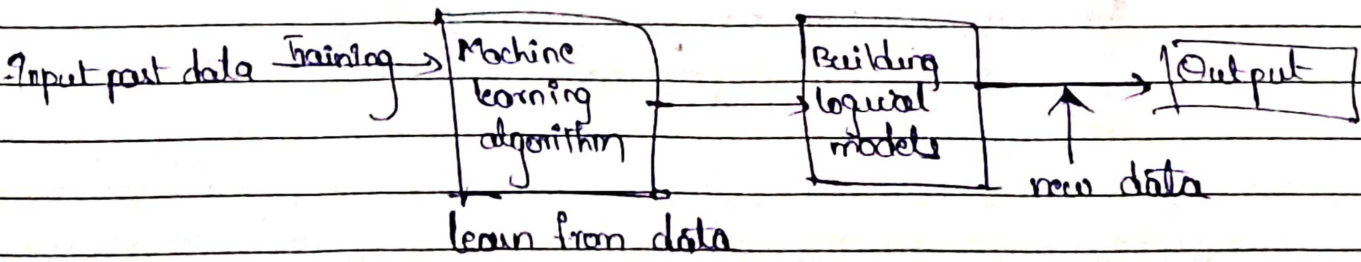
without machine learning
youtube playlist
① we create our own and store them.

machine learning
it creates playlist on it's own without human help.

② flipkart when you search for bed sheets in bought together you get bed cover.

How does ML work?

a machine learning system learns from historical data builds the prediction models and whenever it receives new data predicts the output for it.



Data (Input) → Traditional programming → Output
Program →

Ex:- c, c++, python
f(a,b)
{ return a+b;
}

Data (Input) → Machine learning → Program
~~Program~~ Output →

Ex:-

x	4
2	4
3	6
4	8
6	12

Program
$2 \times \boxed{2} = 4$
$3 \times \boxed{2} = 6$
$4 \times \boxed{2} = 8$
$6 \times \boxed{2} = 12$

Ex:-

x_1	x_2	4
2	1	2
2	2	4
2	3	6

Input Output

$4 \rightarrow \boxed{ML}$ → Program $x_1 \times x_2 = y$

new $x_1 = 2$ $x_2 = 10$ $y = ?$ (20) generate

Classification of ML

Supervised ML Unsupervised ML Reinforcement ML

① Supervised ML :- future predictions, recommendation systems

Ex:- In a locality, there are many house
new house is supposed to construct
we need to predict the price of the house
Broker:- understand the nearly houses price and estimate.
for this we can use machine learning.
Supervised

Unsupervised Machine Learning :- classification task

Ex:- Bucket with lots of fruits of different types.
Based on appearence you divide them and keep seperately like Apples, mangoes, bananas at one place

If we give mixed data, unsupervised learning sees the pattern nature of data and divide them into groups.

Unsupervised learning is all about groupism.

Reinforcement learning:- Day to Day life

Ex:- game, (chess) $\Rightarrow$ computer vs you
Based on your movement game movement gets changed.
automatic game play.

Advantages Of Machine Learning :-

1. Easily identifies trends and patterns.
2. No human intervention needed (automation)
3. Continuous Improvement.
4. Handling multi dimensional and multi-variety data.
5. Wide applications

Disadvantages of machine learning:-

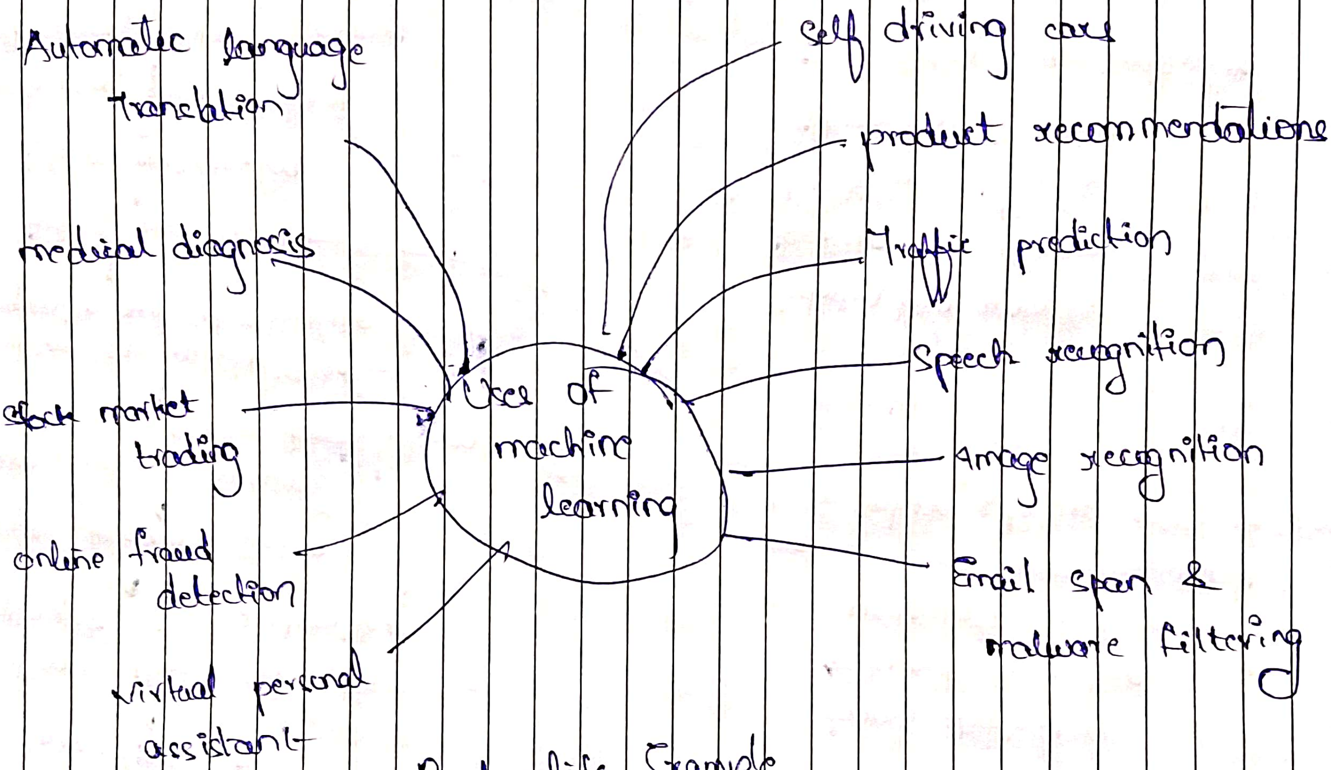
1. Data Acquisition
2. Time and Resources
3. Interpretation of results
4. High error susceptibility.

Ex:- self driving cars

1. Data Acquisition — Needs tons of road data from diff weather, traffic and locations.
2. Time and Resources — Training model takes weeks on high end GPUs and massive storage
3. Interpretation of results — Hard to explain why the car made a specific decision in complex situations.
4. High error susceptibility — a small sensor glitch or rare road scenario can lead to critical mistakes.

a sensor glitch — camera briefly stops working
LIDAR gives wrong distance readings
radar misdetects an object.

even small errors like this can cause the car to make the wrong move like braking suddenly or missing a pedestrian.



Real-life Example

① Camera in laptop ⟶ detect faces.

Complete Roadmap

End to end ML projects

↑

model deployments

↑

Docker / Kubernets

↑

GridSearch, Randomized search,
Hyperopt, Genetics algorithm

↑

Hyper parameter Tuning

↑

linear, Logistic regression,
Decision tree, Random forest, clustering

↑

Machine learning algorithms —
Regression and classification, clustering

↑

Feature selection

↑

Feature engineering

↑

exploratory data analysis

↑

Python — programming language

Correlation
forward elimination
backward elimination
univariate selection
random forest importance
feature selection with decision trees

Exploring data analysis
Handling missing value
Handling outliners
categorical Encoding
Normalizing & standardization

Programming language :- Python, R

R is still great for statistics, Data visualization, research, but python wins when it comes to production - ready machine learning and AI applications.

Exploratory data analysis :

to understand the data it's structure, pattern, relationships, potential problems — before building any models.

why EDA is important ⟶

Imagine trying to cook a new dish — you'd first look at your ingredients understand how fresh they are, and what quantities are available.

EDA is like that — understanding your dataset before cooking up a machine learning model.

EDA helps you.

- Select the right feature for modeling
- Identify patterns that can help in feature engineering
- choose the right model (eg: regression, classification)

Feature Engineering :-

Transforming raw data into meaningful input features that help your model learn better and make accurate predictions.

Real-world example:-

Suppose you are building a health care prediction model that predicts if a patient has certain disease based on input data like:

Name	DOB	Weight (kg)	Height (cm)
Ravi	1995-08-22	75	170

Raw data — not immediately useful
we need to extract useful features from this.

~~aft~~ After feature engineering :-

Age — 28 BMI — 25.95

$$= \frac{\text{weight}}{(\text{height in m})^2}$$

These engineering features are more helpful for the model to learn patterns and make accurate predictions.

Feature Selection :-

Feature selection is the process of choosing the most important input features (columns) that contribute the most to predicting the target variables.

Feature Engineering Vs Feature Selection

Feature Engineering	Feature Selection
① create new from existing	choose best from existing
② raw data → useful data	Removes irrelevant / redundant data
③ Add new column	removes unnecessary columns.

Hyper Parameter Tuning :-

Hyper parameters are the settings you give before training a model and tuning means finding best values for them to get the highest accuracy.

Ex:- Training a decision tree :-
you don't just say train a model,
you also specify how deep the tree should grow –
– That's a hyper parameter.

Some common hyper parameters :

- max_depth → tree grows.
- learning_rate → gradient boosting, neural networks
- n_estimators → no. of trees to use.
- k – for KNN

Docker, kubernets — model to store

What libraries we basically
Use in ML?

① numpy ⟶ numerical problems, mathematical problems.
② matplotlib ⟶ data analysis, graphical format.
seaborn
③ scikit learn ⟶ Ml model
④ SciPy ⟶ constants and functions (like scientific/mathematical constants, eg speed of light, ...) that often need in scientific / technical computations
⑤ Tensorflow ⟶

Deep learning which is a part of machine learning is basically used when you want to work on more advanced tasks.

There are many problems that traditional ml can't handle well and that's when you shift towards deep learning.

For working with deep learning, one of the key tools you'll need is the Tensorflow library — it helps you build and train deep learning models.

Note:- If u're data text / image format, you need 2 additional libraries

- NLTK (Natural language Toolkit): Used for text processing (like understanding and manipulating human language)
- OpenCV — Used for image processing.