INTRO TO AI-ML -1

Data is the new oil.

1. **The need for ML?**

Basic ~~**Data Handling**~~ doesn't require ML. E.g- You have maintained a list of how your conversations with your friend are(Good or Bad) for some 100 days and you will make the decision to lend money to him/her, if there are more Goods than Bads.

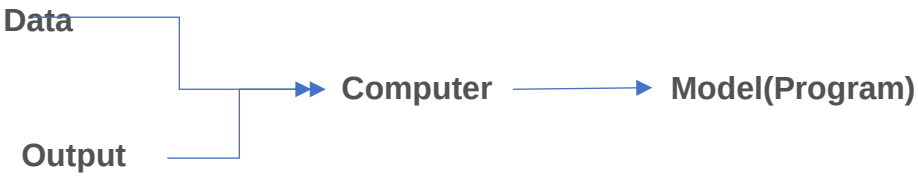
For such purposes, where we derive decisions from data, Excel is enough.

2. **Traditional Programming Paradigm-**

The Developer writes the code(program). The data (inputs) is given to the computer and then the computer computes it in an executable file and the output is generated.

- The output is deterministic(certain).
- E.g - Addition of 2 numbers.

3. **Machine Learning Paradigm-**



- Once the model (brain/concept) is ready, we no longer need the data.
- e.g - If a student is given a Question Bank(Data) and the Answer Key(O/p) and the student reads it and understands the concepts completely, he/she no longer needs the QB and AK and he/she can now solve the problems based on the knowledge gained.
- In this model, the output is predicted . It is not deterministic.
- ML does not require any hand-coded rules or complex flowcharts.
- ML is all about predicting unseen data.

4. **ML Definition** -

ML is all about -

- Improving *performance P*
- *at executing some task T*
- *overtime with Experience E*

e.g- Categorizing the emails as inbox or spam-

- The number (or fraction) of emails correctly classified as spam/not spam -- **P**
- Classifying emails as spam or not spam. -- **T**
- Watching you label emails as spam or not spam -- **E**

5. AI --> ML --> DL

- DL is the subset of ML, which is the subset of AI.
- Domain Knowledge -

e.g -

Walmart USA	India Big Bazaar
It is observed that most of the population which buys diapers also buy beers. --> Parents get stressed due to exhaustive task of handling children hence need beer. So, the store decides to keep the beer stall near to the diapers stall. Thus, the total sale of beer and diapers increased by huge amounts!!	The same data analysis when applied to Indian Store can not be implemented successfully. Instead it will be controversial because Indian parents will never prefer beer for stress reduction due to babies!

Thus, Domain Knowledge is a very important aspect of data science.

6. Supervised learning - Learning in the supervision of someone(teacher).

1. CLASSIFICATION-

- Output is Categorial or discrete in Nature.
- Example - A child has been taught that a shape with 3 sides is a triangle and 0 sides is a circle. So, whenever he is given the no. of sides he can categorize it into a triangle or a circle.
- Machine Implementation - In a machine a graph is plotted based on the data(independent variables). e.g - Sides on x-axis and circular nature on y-axis. Then, the data is plotted and clustered into 2 categories - Triangle and Circle.
- The Machine draws a line between these 2 clusters and calculates the error. The Line such that it equidistantly seperates the 2 categories and has minimum error is found - **Line of Best Fit(LOBF).**
- This Line helps to categorize the data easily.(if the data point lies on the side of triangle, its a triangle or if it's on the side of circle, its a circle.
- Real world Example - Weather Prediction - The model is trained on a dataset of previous days and fed with output labels(Sunny or Rainy) based on the independent variables such as precipitation, humidity, temperature etc. Then, the trained model is fed with new unseen input(data). Based on its learning, it will predict the output as Sunny or Rainy.

2. REGRESSION -

- Output is Continous.
- e.g - Predicting the price of a house based on the sector in which it is situated. Graph is plotted with with Sector no on x-axis and price on the y-axis. Then, **LOBF** is drawn such that it passes through maximum points. Then, when someone wants to know the price he/she can charge for a house in a particular sector, it can be found by finding out the y-coordinate(price) of the line(LBOF), when the x-coordinate(sector no. is known) ---> **Linear regression**

- When there are multiple independent variables on which the dependent variable(price of house) is dependent on such as no. of bedrooms, nearness to markets, availability of gym etc. , the LBOF becomes a **curved graph (polynomial**) - -> **Multiple Regression**