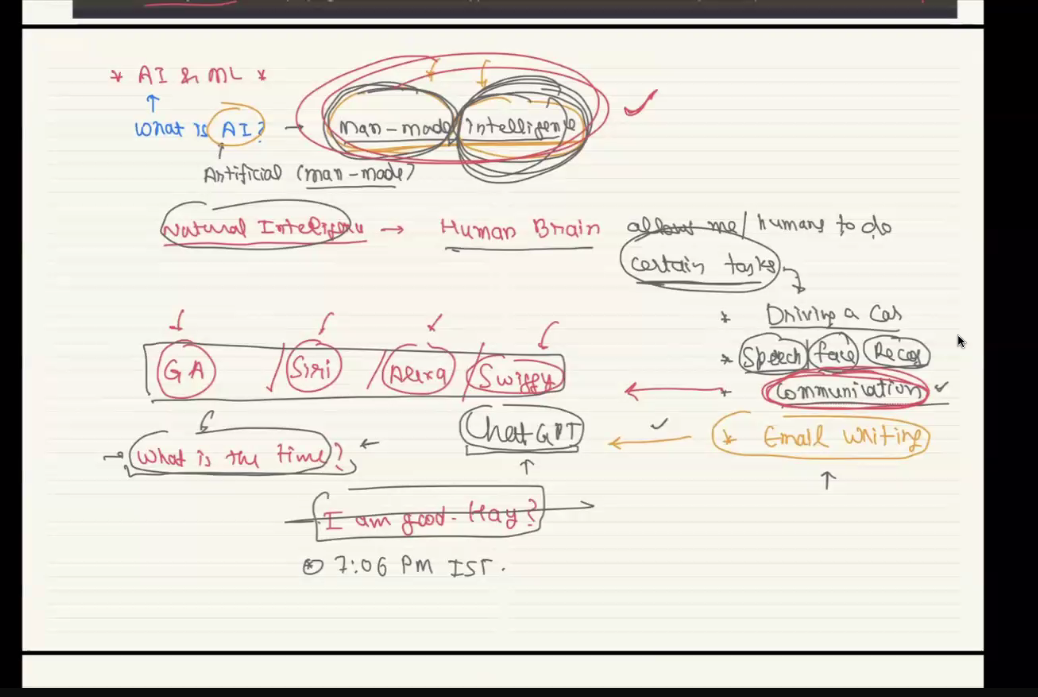
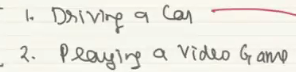
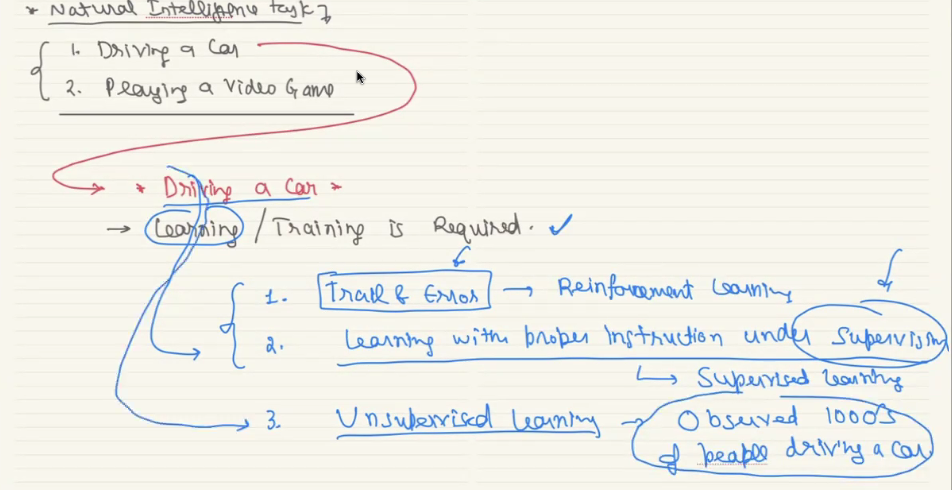
**WHAT IS AI**

****



**WHAT IS NATRUAL INTELLIGENECE TASK**





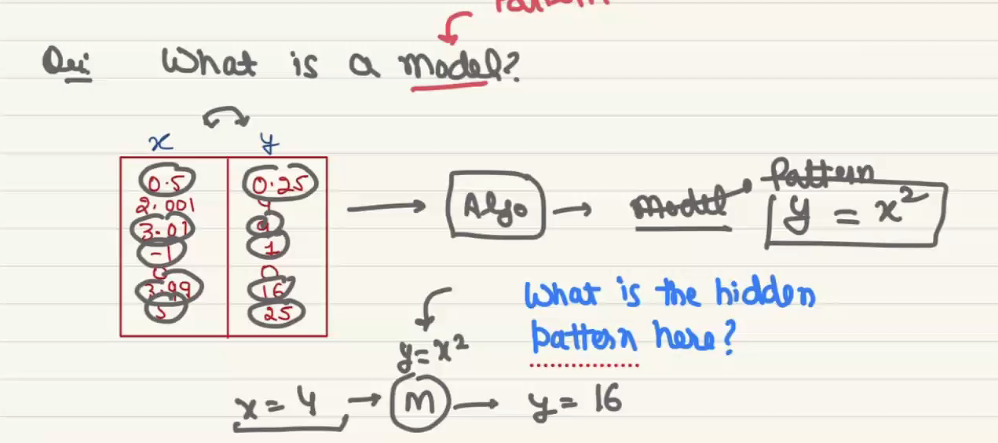
**WHAT IS ML?**

It is process of training machines This is the methods that helps too develop AI human like model.

It is a study of build ml model

**What is model ?**

It is a process finding hidden patterns from pre processed data



In above from column X and Y it will finds the pattern y= x square. This can be done without AI model

But main used of ML/DL is we can build model with very large amount of data

It has three method

1. Supervised
2. Unsupervised
3. Reinforcement

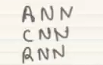
Here machines are learned from data ,patterns and algorithm, when we feed the data with algorithm machines can learn to make prediction. So a model will be developed with algorithm and then data will be feed to that model. Data and algorithm are the two important parts needs to develop ml/dl model, the data will be historical data and the algorithm will be changed as per the data that are going to process.

Deep learning also uses to build model, but it is mainly build a model with very large amount of data while machine learning uses small amount of data, all the model development will be developed from two type of data. Numerical and categorical

**Classification**: A supervised learning task where the goal is to predict discrete labels (categories). Example: Predicting whether an email is **spam** or **not spam**.

**Regression**: A supervised learning task where the goal is to predict continuous values. Example: Predicting **house prices** based on features like size and location.

There are several deep learning model that are:



To build these type of deep learning model we should have good processing unit or hardware that can process very large amount of data.

Some of the hardware are

 **High-Performance GPU/TPU** – NVIDIA GPUs (e.g., RTX 4090, A100) or Google TPUs for fast training.

 **CPU** – A powerful processor (e.g., Intel i9, AMD Ryzen 9) for data preprocessing and model deployment.

 **RAM** – At least 16GB (preferably 32GB+) for handling large datasets.

 **Storage** – SSD (1TB+ recommended) for fast data access.

 **Power Supply & Cooling** – Proper cooling and a reliable PSU for high-performance hardware.

 **Cloud Resources (Optional)** – Platforms like Google Colab, AWS, or Azure for scalable training.

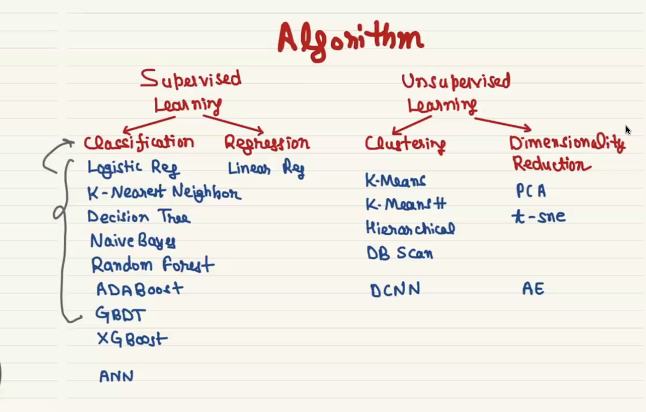
**STEPS IN BUILDING ML/DL**

1. Training - Train model with historical data

We have to do data pre processing, there are lot of pre processing are

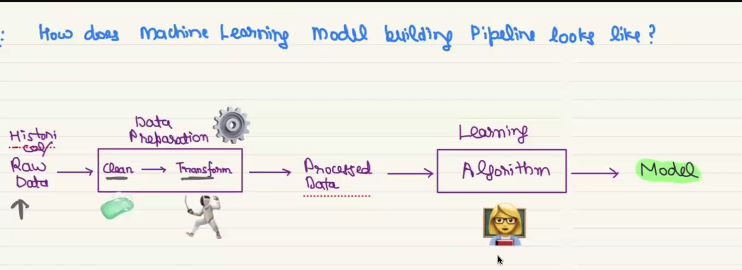
There but it can very form it’s data type

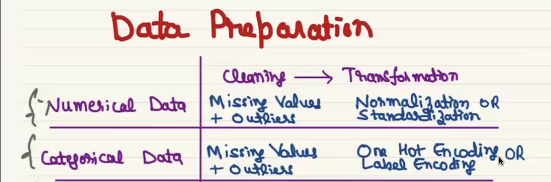
1. Testing of model, evalution phase of mode – Testing the model with test data
2. Inference – Deploy the model for actual prediction of real world data

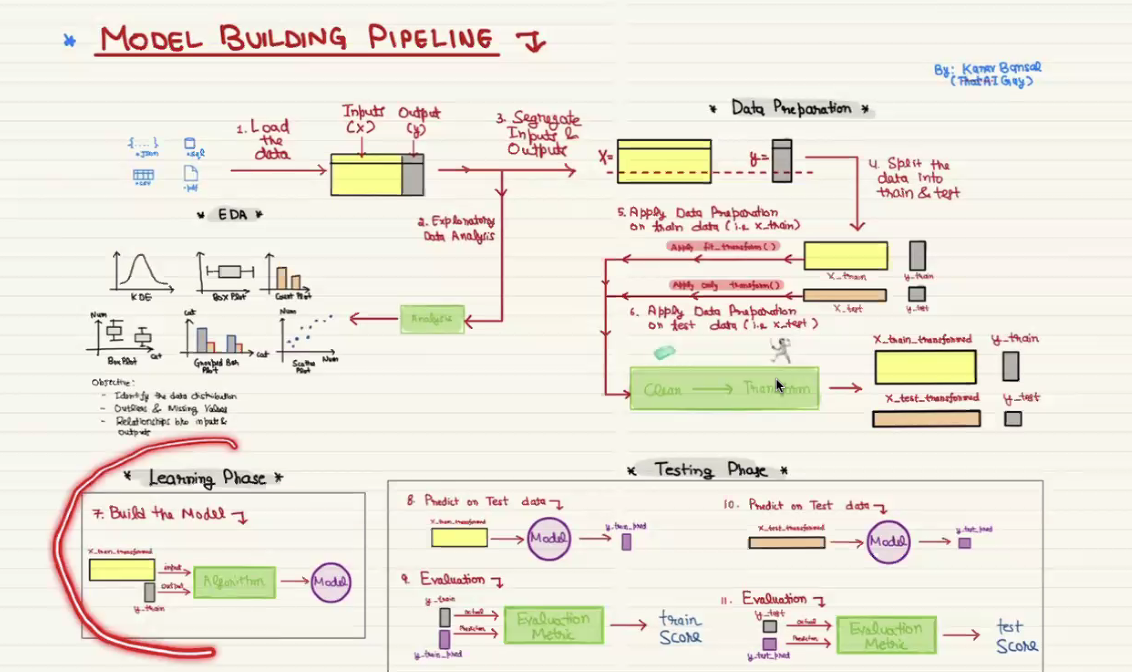


In above ANN, DCNN, AE (auto encoder) are deep learning algorithm used to develop DL model, in auto encoder we changed categorical data to numerical data for the machine under standing.

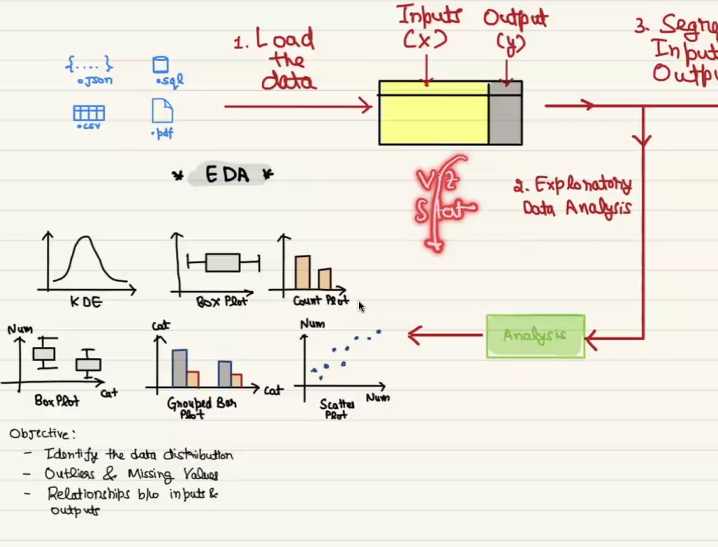
EX: city column has America that will be encoded as 1, and also this is called one hot encoding







**DATA PRE PREOESSING PART**



In EDA we visualize the pre processed data.

In EDA following process will be performed to make visualization.

In **Exploratory Data Analysis (EDA)** in Machine Learning, variables can be analyzed based on their relationships and the number of variables involved:

**1. Univariate Analysis (One Variable) 📊**

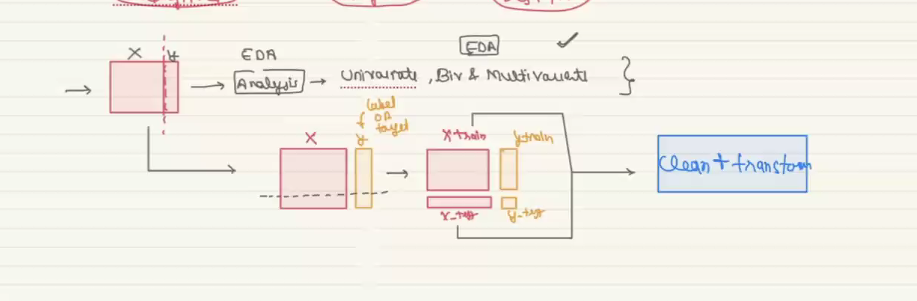
* **Definition**: Analysis of a **single variable** without considering relationships with other variables.
* **Example**: Analyzing the **height distribution** of students using histograms or box plots.
* **Techniques**: Histograms, Box plots, Summary statistics (mean, median, mode).

**2. Bivariate Analysis (Two Variables) 🔄**

* **Definition**: Analyzing the **relationship between two variables**.
* **Example**: Examining the correlation between **study time** and **exam scores**.
* **Techniques**: Scatter plots, Correlation coefficient (Pearson/Spearman), Regression analysis.

**3. Multivariate Analysis (More than Two Variables) 📈**

* **Definition**: Studying the relationships among **three or more** variables.
* **Example**: Analyzing how **age, income, and spending habits** influence each other.
* **Techniques**: Heatmaps, Pair plots, PCA (Principal Component Analysis), Multiple Regression.



Data preprocessing should be done after train test split, if we do it before then **data leakage** problem occur.

In evaluation phase the model will be tested with the model it uses on train phase, on deployed platform only model will have real world data. Model will be deployed when the accuracy of evaluation phase is good, this can be called as train score.

**WHAT IS DATA LEAKAGE ?**

It is like question paper leakage for exam, like this in model training phase we given the data which is used to test the model performance, because of this reason we have to do data preprocessing after the train test split, data preprocessing will be done on whole data, so when we give the whole data on training the question paper leakage happens.

Candel stick plot on stock market open and close is the time f stock market timing of open and clode



In second question we want to predict house price so it is unsupervised learning, the target label is house price prediction, but in data have only the house details like below

