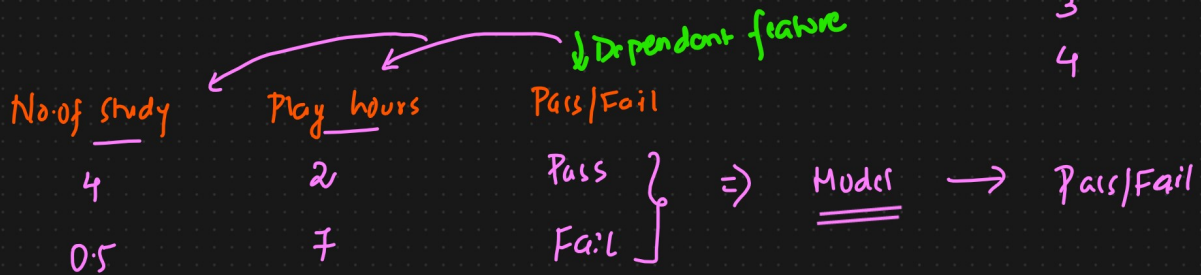
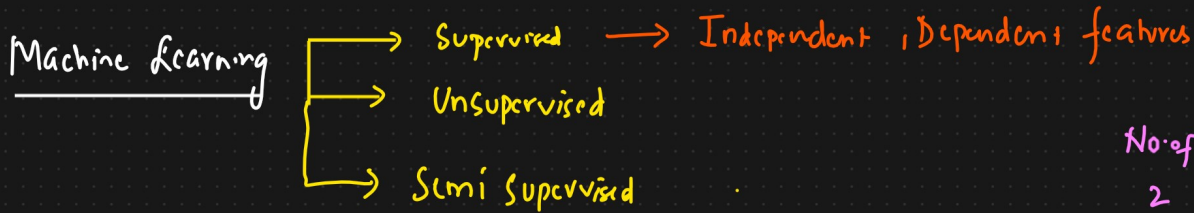


# Entire Development of NLP



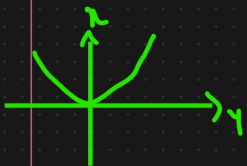
⇓  
ANN

⇓  
Tabular Data → features

- Numerical
- Categorical

No. of children family  
2  
3  
4

Artificial Neural N/w



Machine Learning Models

ANN Models

1980's

⇒ Numerical calculation

Text

30 km

↓  
language changes

$$y = x^2$$

Research      ÷      2014      ⇒      One hot Encoding

⇒ Text → Numerical values ⇒ Vectors ← DATA

10 years

Models ⇒ Arithmetic Equations

Models

Naive Bayes

RNN, LSTM RNN, Bidirectional LSTM RNN,

Encoder Decoder, Transformers, BERT

TEXT  $\rightarrow$  Numerical Values  $\Rightarrow$  Text Embedding or Word Embedding.

① OHE  $\Rightarrow$  One Hot Encoding.

Vector Size  $\Leftarrow$  10,000 Unique Words

[ KRISH IS A GOOD TEACHER ]  $\rightarrow$  [ [1 0 0 0 0], [0 1 0 0 0],  
[ 0 0 1 0 0 ], [0 0 0 1 0]  
[ 0 0 0 0 1 ] ]

② BoW [Bag of Words] . ③ TF-IDF [Term Frequency - Inverse Document Frequency].

Deep learning Model  $\div$  Text Embedding Models

① Word2Vec [ ]  $\Rightarrow$  ANN  $\Rightarrow$  Glove  $\Rightarrow$

King - Man + Queen = Women  $\leftarrow$

② Avg Word2Vec = Text  $\Rightarrow$  Vectors  $\Rightarrow$  Numerical Representation

RNN, LSTM RNN, GRU, Encoder Decoder Huggingface

Attention Is All You NEED  $\Rightarrow$  TRANSFORMERS  $\Rightarrow$  Open Source

BERT  $\Rightarrow$  GPT  $\Leftarrow$  Open AI  $\Rightarrow$  Generative Pretrained Transformers  
 $\Downarrow$  Google Gemini Pro PALM  $\rightarrow$  BARD  $\rightarrow$  Google Gemini

TRANSFORMER  $\Rightarrow$  Encoder Decoder }  
BERT  $\Rightarrow$  Encoder

Vector Embedding  $\Rightarrow$  Huge DATASET  $\xrightarrow{10,000 \text{ documents} \Rightarrow \text{Text DATA.}}$  Text DATA  $\rightarrow$  Vectors

$\downarrow$   $\downarrow$

Word or Text  $\rightarrow$  Vectors { [ ]  
[ ]

Startup  $\Rightarrow$  Vector DATABASE

CHROMADB, FAISS, ASTRODB, Pinecone

$\downarrow$   
RAG { Retrieval  
Augmented  
Generation }

Support  $\Rightarrow$  RAG

KRISHNAIK.IN

YT videos  
Github  
Courses

$\leftarrow$  Pinecone DB  $\rightarrow$  Open Lim Group chat

CHATBOT  $\Rightarrow$  What is ML?

{ - - - - }

What is ML.



Agentic RAG, Corrective RAG

KG graph  $\Rightarrow$  Knowledge graph

hLMops  $\Rightarrow$  AWS  $\Rightarrow$  Sagemaker  $\Rightarrow$