Natural Language Processing

Textual data

→ Linguistics that enabled the Computers to understand, interpret & generate the human language.

① Google Translation

I love data science

↓

मुझे डेटा साइंस पसंद है

② Sentiment Analysis

Today's movie is quite boring. I have just wasted two hours watching this movie.

↓

Negative

③ Language Models → Predict next word

↳ Autocompletion tasks (Email)

I hope you are doing well.

Code → further depending upon function name it gives all the further code.

(4) Text - to - image

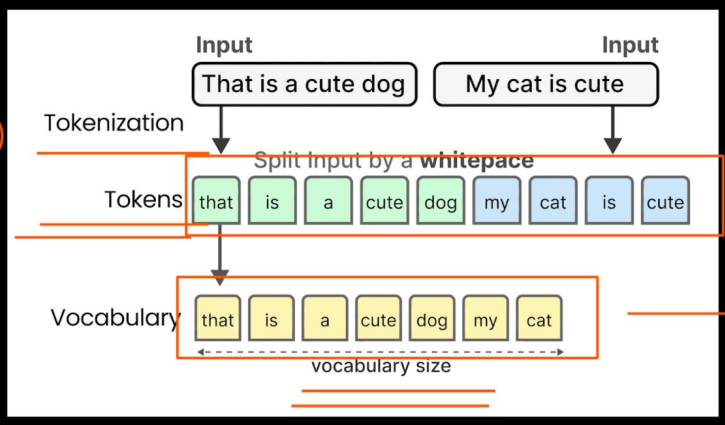
Step 1 → Computers/machines will be able to understand the textual data.

Embeddings → textual data

↓

numeric vector

Textual Data Preprocessing

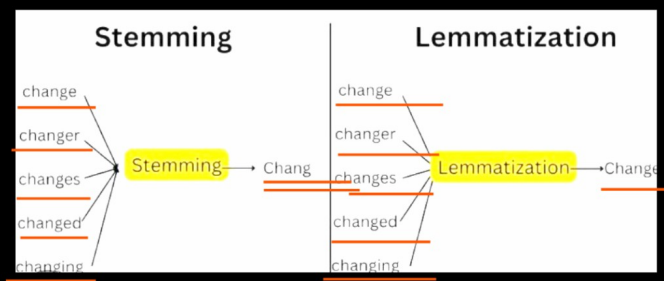


1. Tokenization → lowercase
   - Vocabulary → 7-words

2. Stopword removal (a, an, the, is)

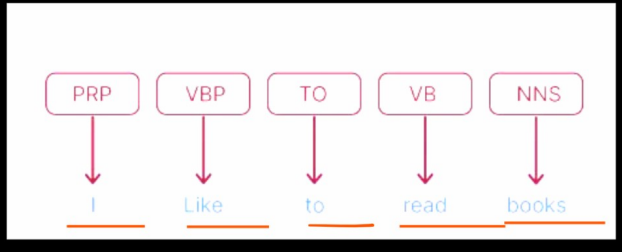
3. Stemming & Lemmatization (Root word)

[run, ran, running] → run



↳ has a literal meaning

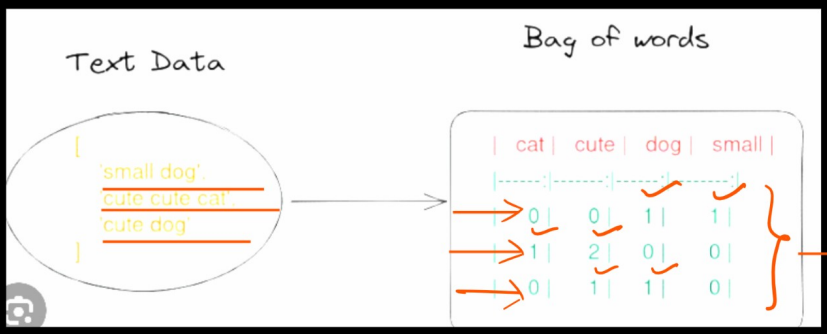
4. POS Tagging



5 Embedding → to the numeric vectors

Bag of words → no context (frequency based)

small, dog, cute, cat → unique words



→ frequency of the individual words

Word2vec → Static embeddings

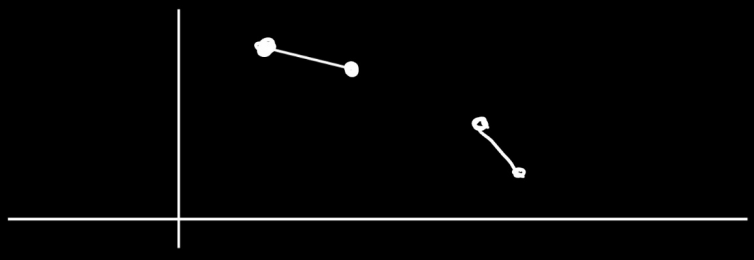
	cats	puppy	houses	apple	baby
animal	.91	.93	-.56	-.67	.01
newborn	-.11	.71	-.32	-.1	.90
human	.19	.36	.31	.29	.87
⋮	⋮	⋮	⋮	⋮	⋮
plural	.94	-.82	.94	-.51	-.11
fruit	-.51	-.91	-.5	.89	-.51

Dimensions

Number of dimensions (properties)

1024 values!

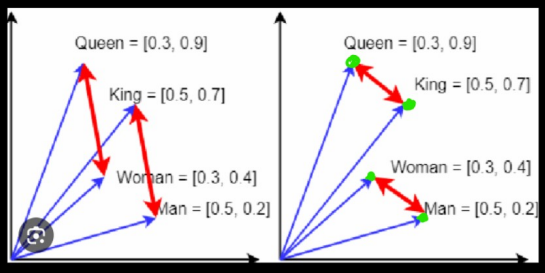
Pretrained model



King + Queen − Man

↓

Vector



Languages → ambiguous

- apple (I am eating fruit)
- apple (I am currently working in company apple)

Word2vec → Single vector

↓

{ Dynamic embeddings }

↑

RNNs

↳ Attention based models → Transformers

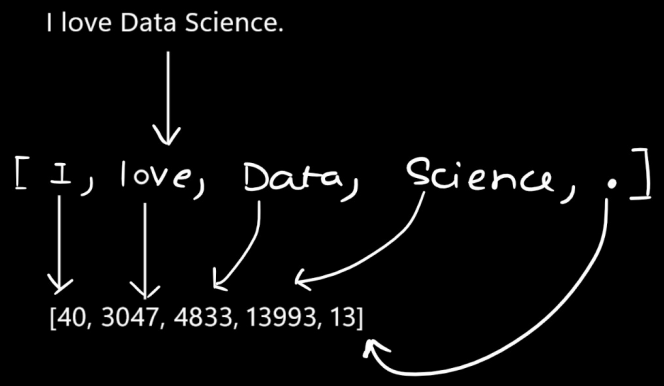
↓

GPT → Generative Pretrained Transformers

Attention is all you need

Pretrained Models → transformers

1000 Records ← finetuning → Customize the above pretrained model for your dataset.



Task + ML Project Ready

- credits → $5
- secret key → create
- .env → define
- tokenization, embedding → execute
- deep analysis → dimensions &

What each dimension represent