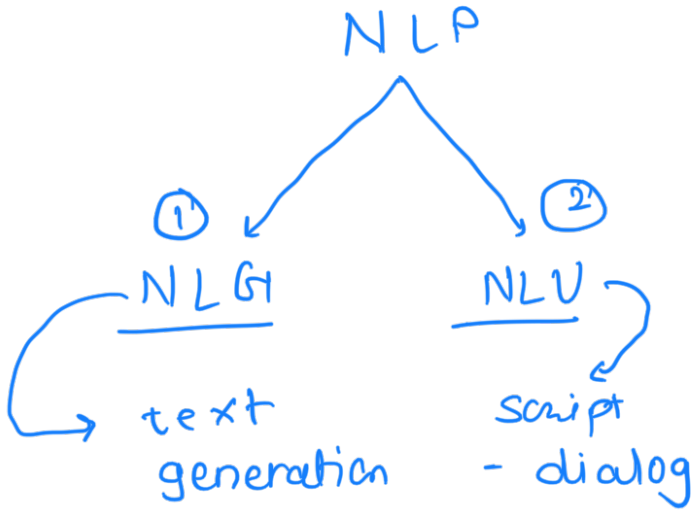


Basic Terms

1. Phonemes - character, sound - smallest unit
2. Morphomes & lexemes → words
3. Syntax → sentence, phrase.
4. Context → meaning

lexemes { learning, learn, learned } learn

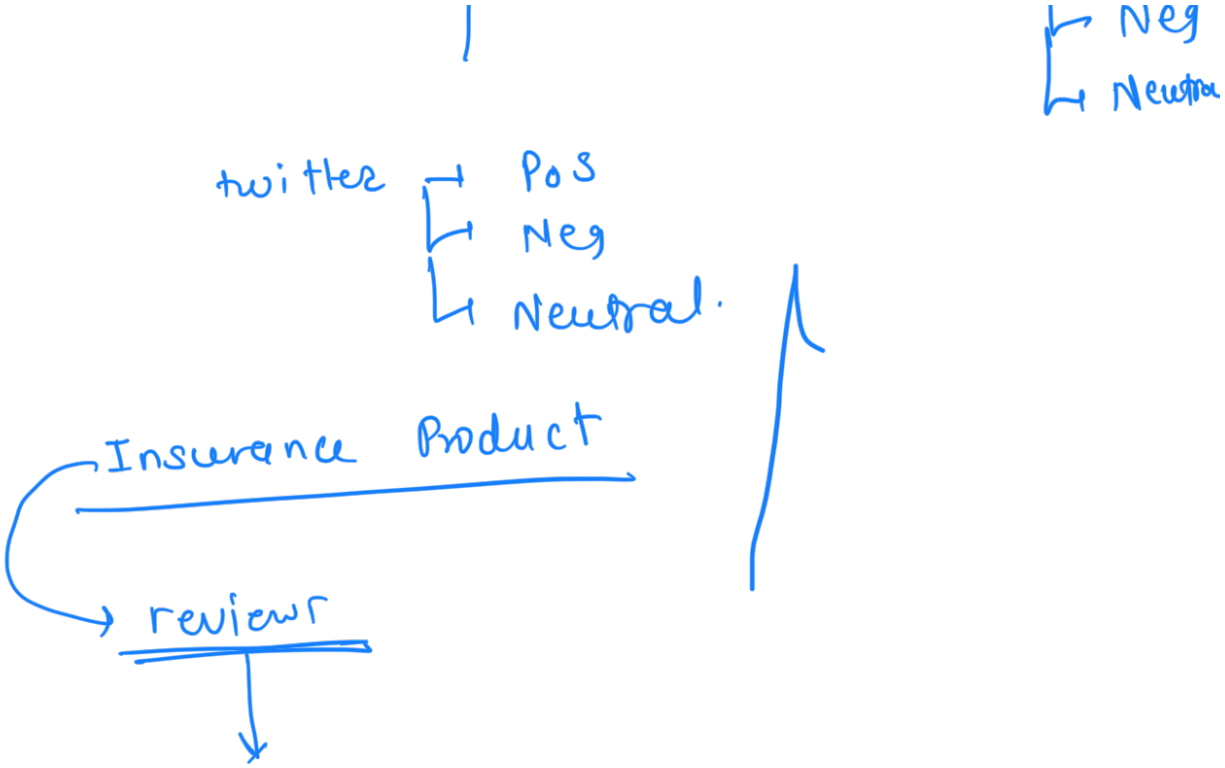
Applications

1. Sentiment Analysis → classification

text	target
review1	0
review2	1

0, 1, 2

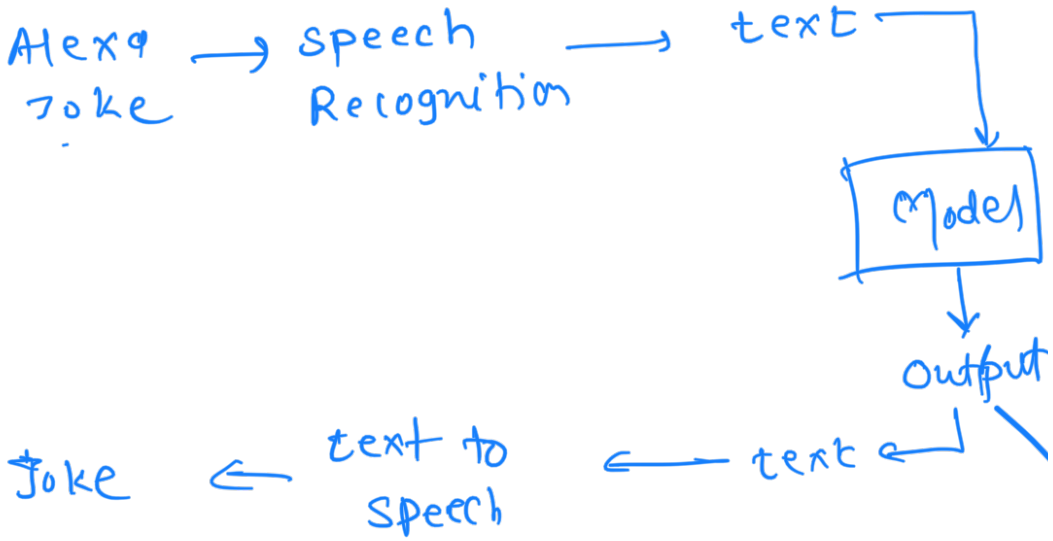
review → pos



② Document classification

text	target
Doc 1	Aadhar
Doc 2	D L.
Doc 3	I.P.

③ Google Assistant.



④ chatbot → NLU

- Greeting Hi → Hello
- → order (1234) → ≡

⑤ Text Summarization

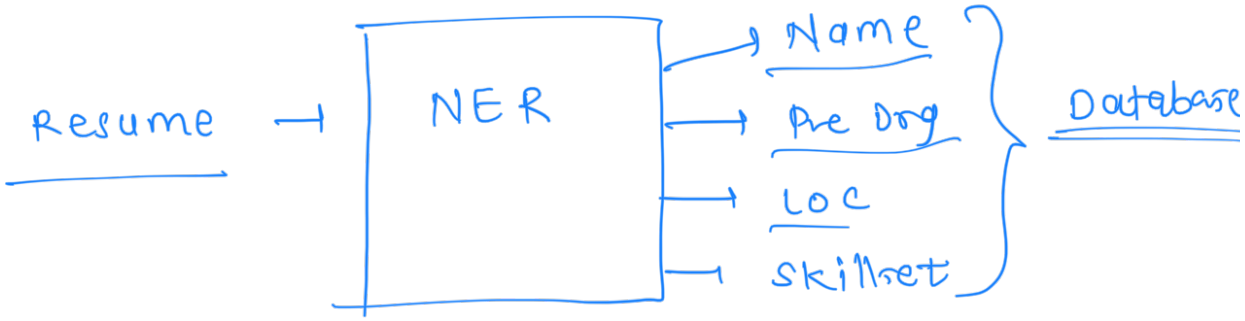
1000 doc → 10,15

→ ① Extractive → NLU

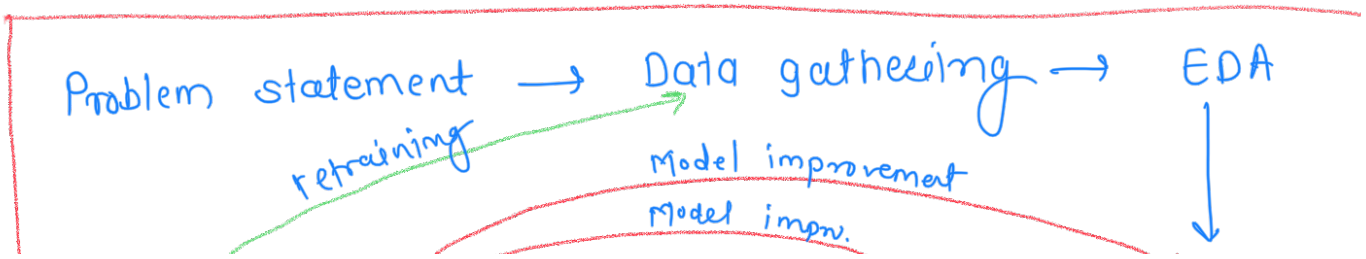
→ ② Abstractive → NLG

⑥ Named Entity Recognition

- Resume Parsing



NLP Pipeline



Deployment ← Evalu. ← Modelling ← Feature Engineering ← Preprocessi

NLP Roadmap

1. Preprocessing
   - Tokenization
   - stemming / lemmatization

2. Text to numeric
   - → Count vectorizer
   - → TFIDF
   - → word2vec
   - → DOC2VEC

3. M.L. → classification

4. Deep learning
   - RNN
   - LSTM
   - GRU

   AI

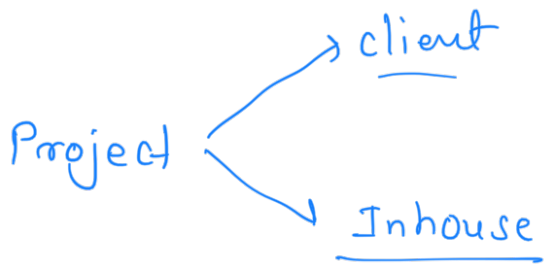
5. Model Building with deep learning

6. Transformer
   - Attention

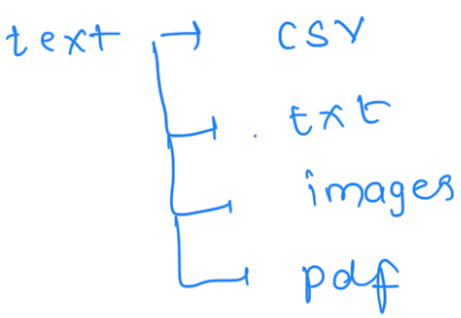
- Attention
- Self Attention
- Multi Head Attention
- encoder & decoder

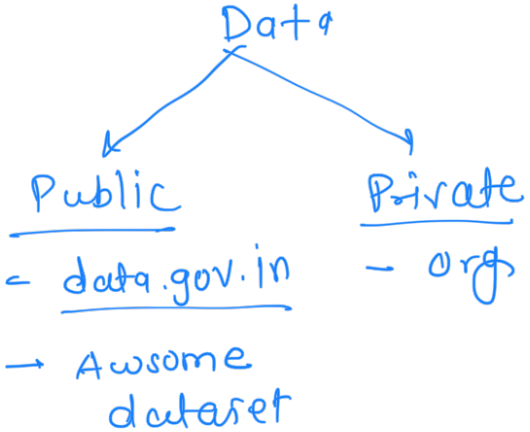
7. BERT
- GPT 3
- T5

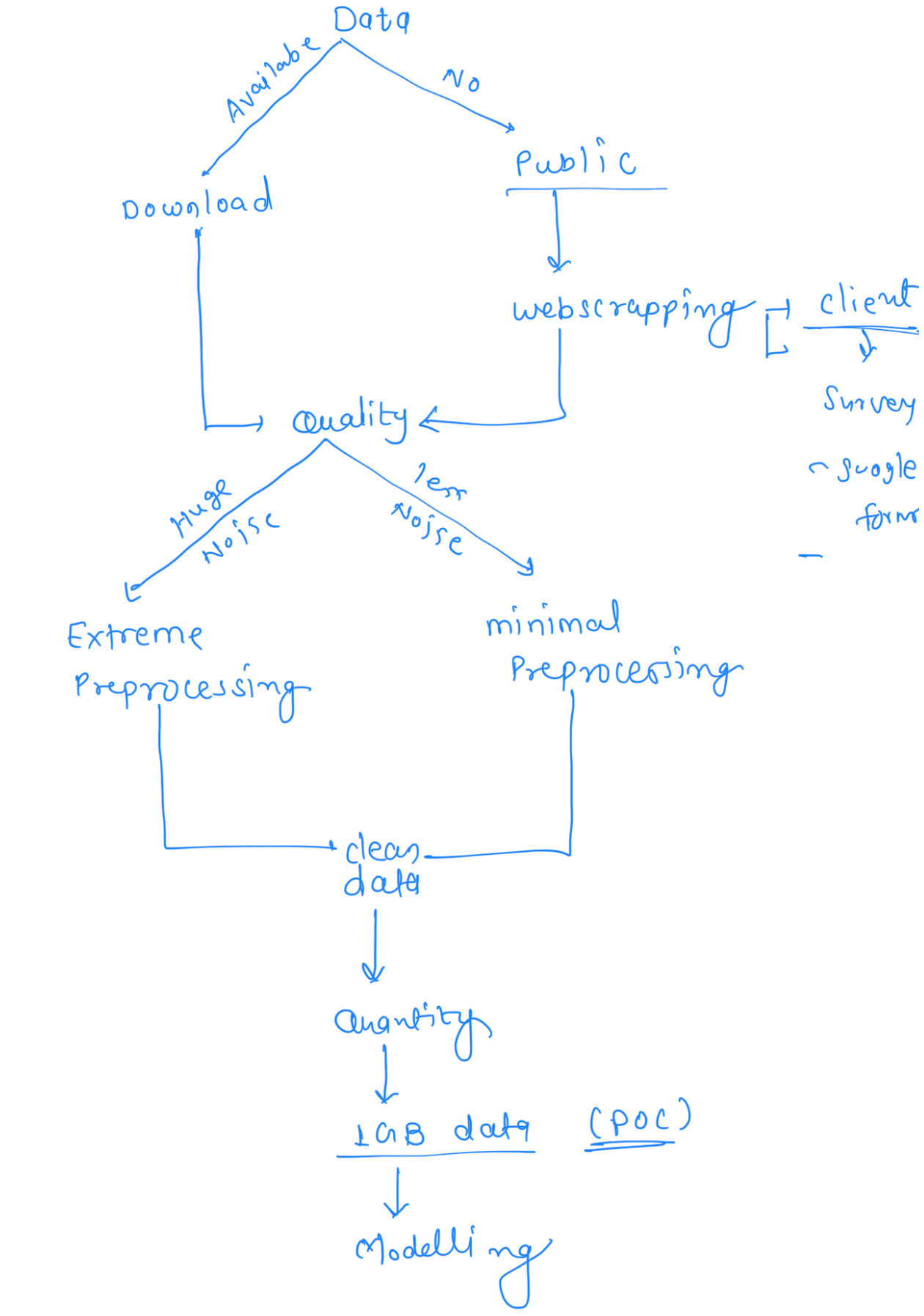
1. Data Gathering



- website →
- database →
→ api

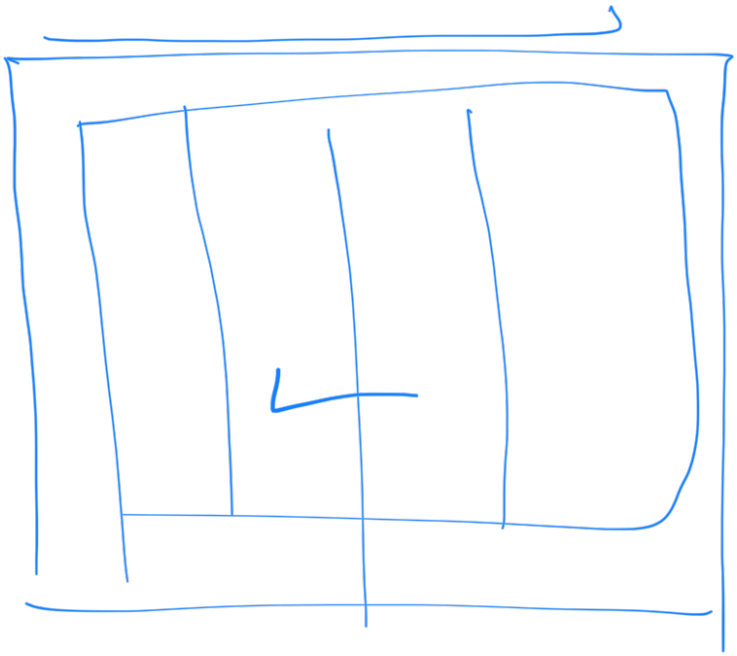





Data
Available
No
Download
Public
webscrapping
client
Survey
- Google
form
Quality
Huge
Noise
less
Noise
Extreme
Preprocessing
minimal
Preprocessing
clean
data
Quantity
1GB data
(POC)
Modelling


client
database
SQL
Queries
cloud

Scheduler
- 1 day
→ MLOps
→ automation



- Beautiful.

(2) EDA

(1) Ngram

(2) Word cloud

(3) Keyphrase Analysis

(1) Ngram

"Minakshi is working in google"

(i) Unigram

- ["Minakshi", 'is', 'working', 'in', "google"]

(ii) Bigram

- [" Minakshi is", "is working", "working in", "in google"]

(iii) Trigram

-

(1) POS → words analysis

Neg →

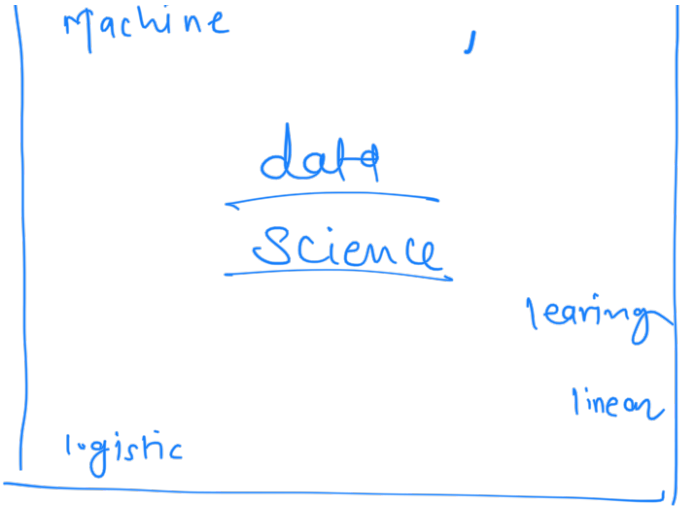
(2) extract stopwords

review① = We like the product

review② - Awsome product, I will recomm to everyone

review③ - Great customer service, staff is quite polite.

(2) Word cloud



⑧ keyphrase andlysis

- We are learning nlp

✱ Preprocessing

① Tokenization

① Sentence Tokenizer → Punctuation, Syntax, Conjunction

② Word Tokenizer

- ① white space Tokenizer
- ② word Tokenizer

③ Regex Tokenizer

② Normalization → lowercase

Great → num1

great → num2

Removal

③ Stopwords

" Arun is suffering from a cancer. doctor Manish is treating him. Nurse Rutika just gave capsule XYZ "

Healthcare
— Nurse, doctor, patient, bed, appointment,

Corpus → collection of words, sentence

Corpora → collection of corpus