* NLP Pipeline

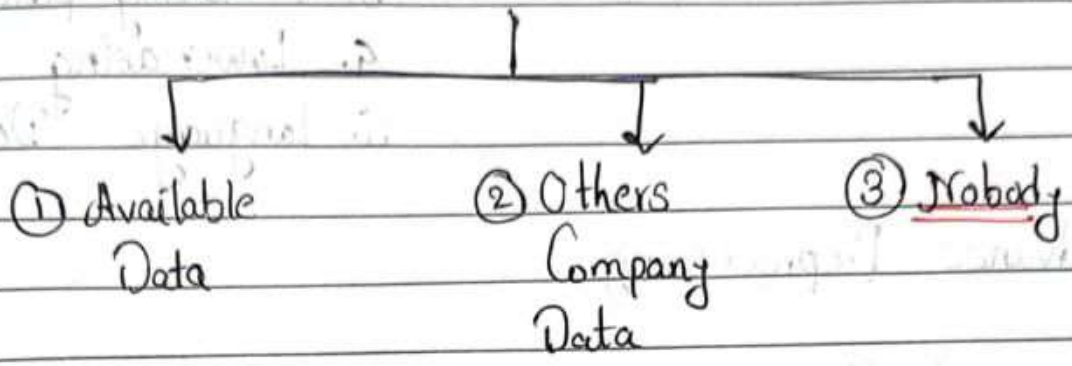
Set of steps followed to build an end to end NLP Software.

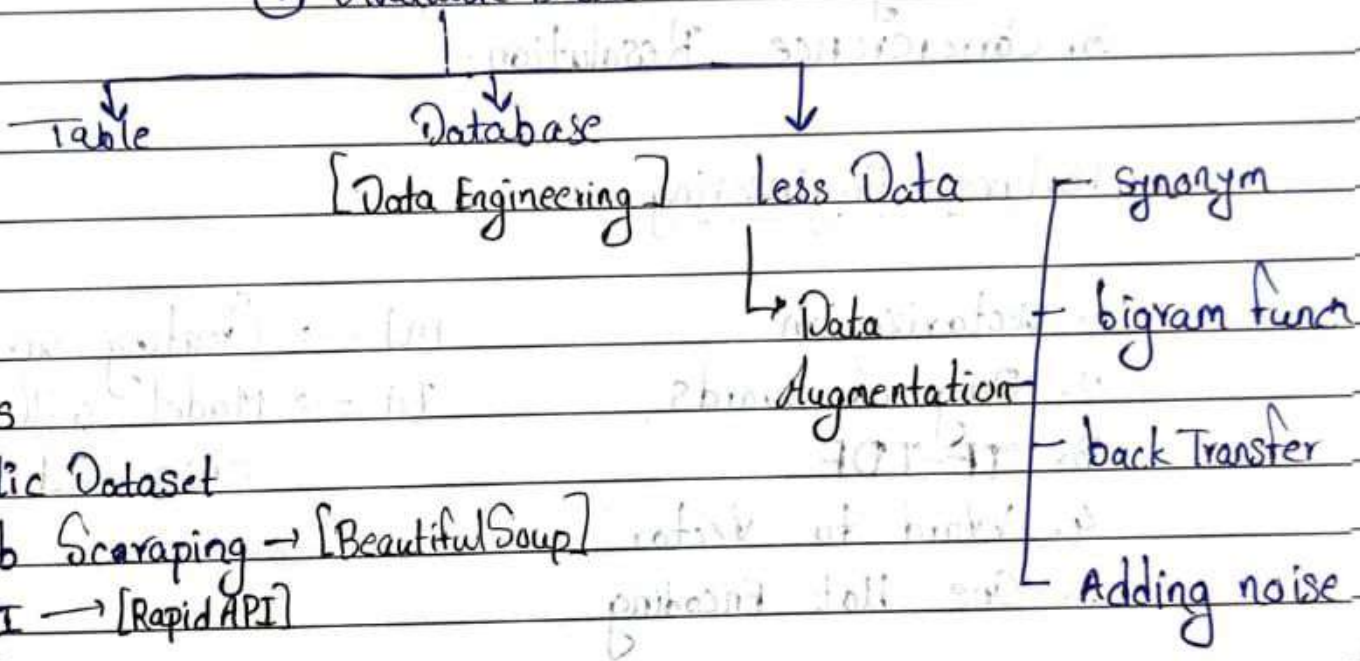
* Steps:

1. Data Acquisition
2. Text Preparation
   a. Text Cleanup
   b. Basic Preprocessing
   c. Advance Preprocessing
3. Feature Engineering
4. Modelling
   a. model building
   b. Evaluation.
5. Deployment
   a. Deployment
   b. Monitoring.
   c. Model Update.

→ Not universal pipeline.
→ Deep learning Pipelines are Slightly different.
→ Pipeline is non-linear.

1. Data Acquisition:





② Others
- → Public Dataset
- → Web Scraping → [BeautifulSoup]
- → API → [RapidAPI]
- → PDF
- → Image
- → Audio → [Speech To text]

2. Text Preparation:

a. Cleaning
- i) Html Tag cleaning
- ii) Unicod Normalization (emoji)
- iii) Spelling checker

b. Basic Preprocessing

1. Basic:
   - Tokenization → Sentence
   - Tokenization → Word

2. Optional
   1. Stop Word removal
   2. Stemming / lemmitization
   3. Removing punctuations, digits
   4. Lowercasing
   5. Language Detection.

c. Advance Preprocessing

1. Part of Speech Tagging (POS)
2. Parsing
3. Correference Resolution.

3. Feature Engineering

1. Vectorization
2. Bag of Words
3. TF-IDF
4. Word to Vector
5. One Hot Encoding

ML → Creating our own features
DL → Model will create its own feature Embeddings

4. Modelling

1. Applying models
   - → Amount of Data
   - → Nature of Problems

   { a. Heuristic Approach
   b. ML
   c. DL (Transfer learning)
   d. Cloud API

2. Evaluation
   - a. Intrinsic evaluation (Confusion matrix, Accuracy matrix)
   - b. Extrinsic evaluation. (Evaluation After Deploying product)

5. Deployment:

A. Deployment
→ API (microservice)
→ Chatbot

B. Monitoring
→ Creating Dashboard with various evaluation matrix [webpages Dashboarding]

C. Update
→ Updating on Server.