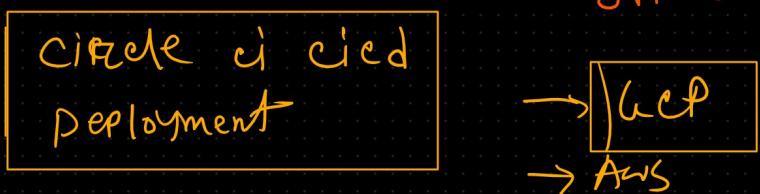


End to End NLP project Implementation

Project Overview:

- * Introduction of project
- * Problem statement
- * Solution understanding
- * Code walkthrough → Notebook Experiment
- * Deployment part → cical AWS deployment
with auto scaling



* Text classification

NER
Text Summarization
Sentiment Analysis

Flask
Streamlit



RNN →

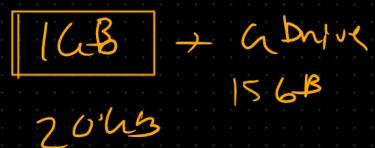
Hate speech classification

Hate speech classification is the process of identifying and categorizing speech or language that may be considered offensive and harmful.



① Notebooks Walkthrough

② Architecture Diagram



③ cost involved,

→ cloud bucket → S3 bucket

→ Train model — CPU — X

→ model push — cloud bucket

→ Deployment —

Prerequisites:

① Basic of NLP → RNN → Text classification

② Python Programming — OOPS

③ Your Dedication

④ CCP, AWS

End to End project

NLP project → NLP pipeline

Breakdown the problem into several sub problems.

① Data acquisition

② Text preparation

 |
 | → Basic preprocessing
 | → Advance preprocessing

③ feature Engineering

④ modeling

⑤ Evaluation

⑥ Deployment

⑦ monitoring and model update

① Data Acquisition:

→ Available Data (CSV, TXT, PDF)

→ Others Data (DB, Internet, API, Scrap)

→ No Data (create your own data) - LM

Note! less data (augmentation)

→ replace with synonyms

Example:

I am a Data scientist

I am a AI Engineer

→ Biagram flip:

① I am Bappy

② Bappy is my name

→ Back Translate:

→ Adding Additional noise

I am a Data scientist, I love this job

(2) Text Preparation:

① Basic preprocessing

② Advancee preprocessing

① Basic:

① Tokenization [sentence] word

other step:

plays, played, playing

- ① stop word remove
play
- ② stemming / Lemmatization
- ③ removing punctuation (., ; : # ?)
- ④ lower case → I am Bappy, bappy
is my to army alias name

② Advance Preprocessing'

- ① parts of speech tag (POS)
- ② parsing
- ③ confection resolution

③ Feature Engineering

→ Text Vectorization — Text to vector/number

→ TFIDF, Bag of word, One hot, Word2Vec.

Embeddings model

→ Encoding

hate	— 1
no hate	— 0

④ modeling:

- ① Heuristic ~~X~~
- ② ML ~~X~~
- ③ DL - RNN ✓
- ④ Cloud API



⑤ Evaluation:

① Intrinsic

② extrinsic

accuracy score, F1, precision, recall

→ Production

⑥ Deployment: