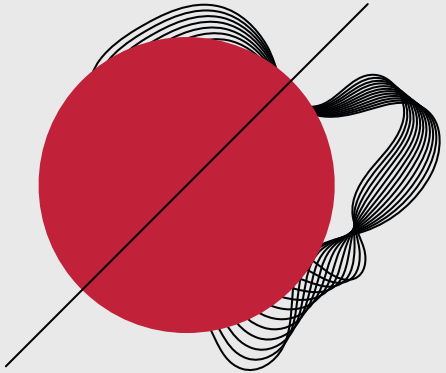
COVID-19 VACCINE RESPONSE



CZ4034 Information Retrieval Group 24

OUR TEAM

Manav Arora

Chengxuan

Lu Cheng

Akshaya Muthu

Parth Taneja

TABLE OF CONTENTS

01. **INTRODUCTION**
Why analyse COVID-19 vaccine response?

02. **CRAWLING**
How we got our data

03. **INDEXING & QUERYING**
UI for Querying

04. **CLASSIFICATION**
Sentiment prediction for COVID-19 Vaccine

01.

INTRODUCTION

131M

Number of Cases

2.85M

Number of Deaths

2%

Percent of people fully vacinated

MIXED SENTIMENTS TOWARDS VACCINE

02.

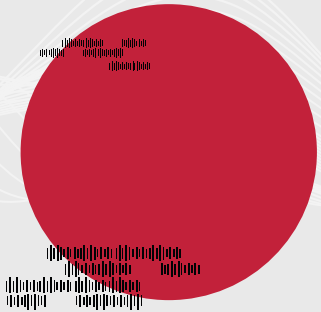
CRAWLING

60,775 TWEETS CRAWLED

Keywords:
covid vaccine, coronavirus vaccine, corona vaccine

INFORMATION RETRIEVABLE

- Total number of **Positive, Neutral & Negative** Response
- Tweets containing specific **keyword**
- Tweets from a specific geographic **location**
- Specific **precautionary measures and safety protocols** with respect to the vaccine
- **Approval, effectiveness, access, updates** regarding to vaccine



Records, Words and Types

Total Tweets	Total Number of words		Total Number of Unique Words	
	Pre-Processing	Post-Processing	Pre-Processing	Post-Processing
60,775	1,453,183	772,001	159,051	30,237

03.
INDEXING & QUERYING

Indexing

INDEX

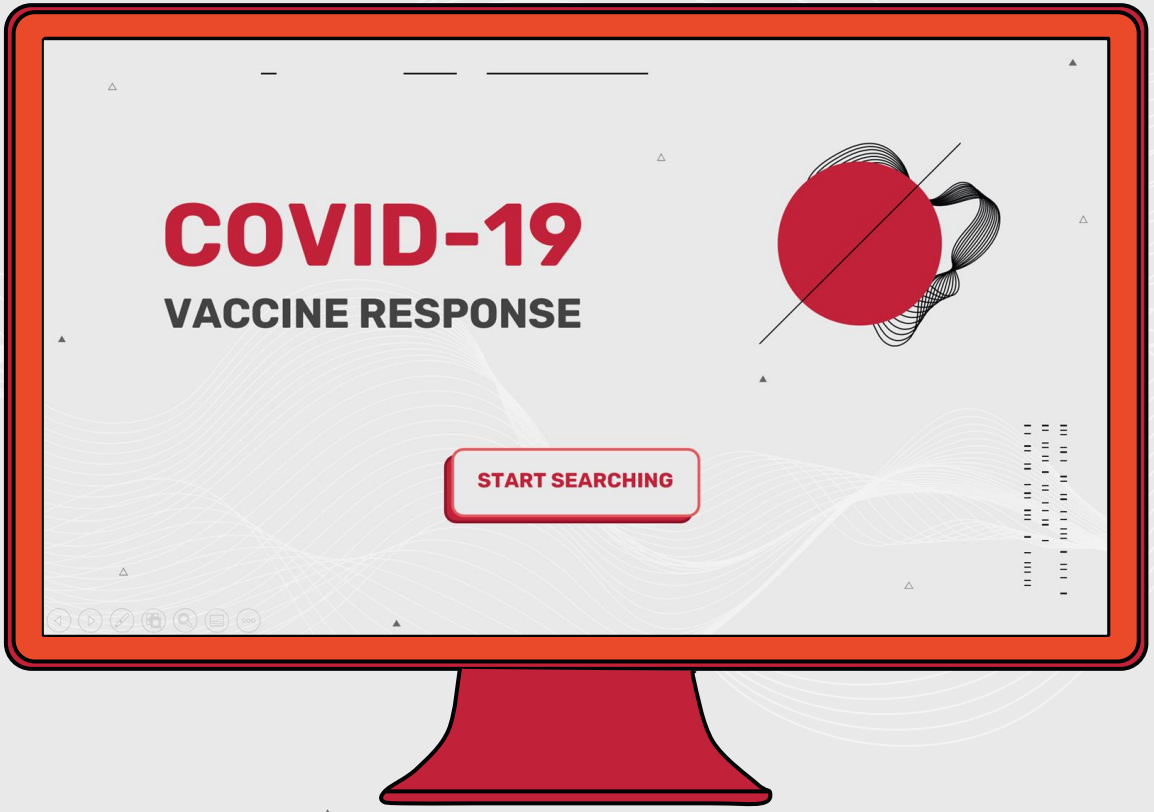
Apache SOLR 8.8.1 is used to index the crawled tweets

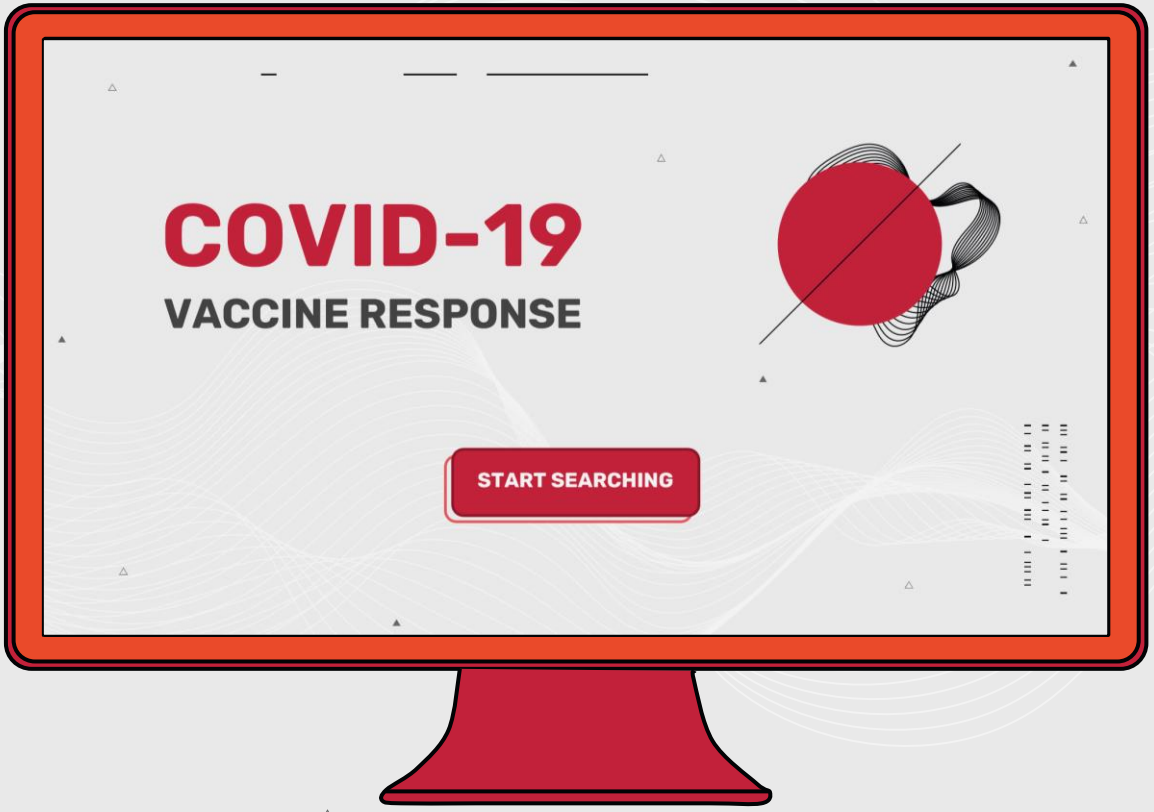
QUERY

JSON Result returned via HTTP GET

WEB INTERFACE

Django is used to built the web interface that displays the query result


COVID-19
VACCINE RESPONSE
START SEARCHING


COVID-19
VACCINE RESPONSE
START SEARCHING

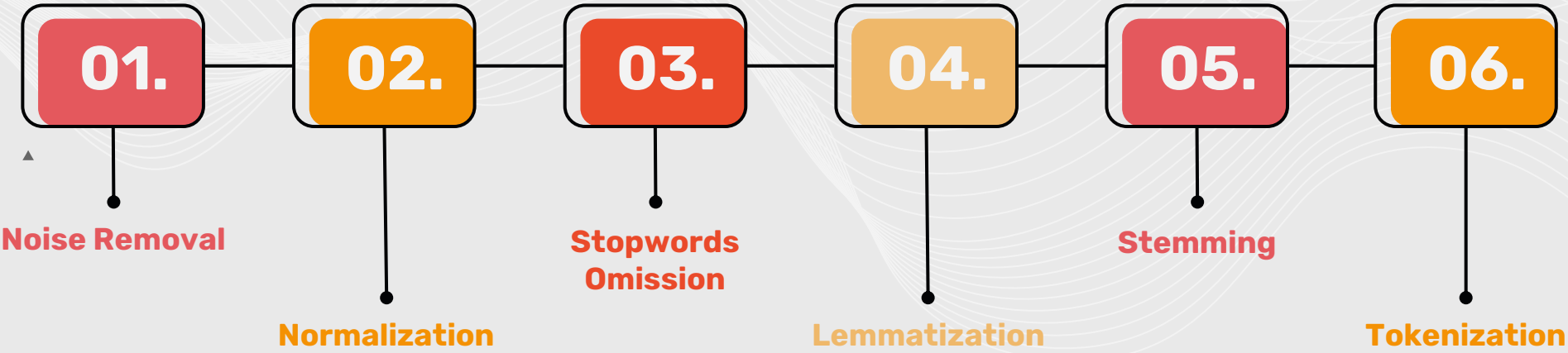
Speed of Querying

Query	Speed of Querying (milliseconds)
Processed:hate	0 ms
Processed:”pfizer vaccine”	26 ms
Processed:"worry" OR "hate"	6 ms
Processed:angry^4 covid	1 ms
Processed:happy vaccine~10	1 ms

04.

CLASSIFICATION

Pre-Processing



Results of Comparison Models

Model	Count Vectorizer			Tf-idf Vectorizer		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Naive Bayes	0.824	0.833	0.824	0.812	0.800	0.800
KNN	0.788	0.735	0.735	0.790	0.757	0.757
SVM	**0.944**	**0.944**	**0.944**	**0.953**	**0.953**	**0.953**
Decision Tree	0.801	0.798	0.797	0.790	0.790	0.790

Proposed Model

Bi-LSTM

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 182, 50)           1228350
_________________________________________________________________
dropout (Dropout)            (None, 182, 50)           0
_________________________________________________________________
bidirectional (Bidirectional (None, 128)               58880
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 3)                 387
=================================================================
Total params: 1,287,617
Trainable params: 59,267
Non-trainable params: 1,228,350
_________________________________________________________________
```

BERT

```
Model: "tf_bert_for_sequence_classification"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
bert (TFBertMainLayer)       multiple                  109482240
_________________________________________________________________
dropout_37 (Dropout)         multiple                  0
_________________________________________________________________
classifier (Dense)           multiple                  3845
=================================================================
Total params: 109,486,085
Trainable params: 109,486,085
Non-trainable params: 0
_________________________________________________________________
```

Results of Proposed Model

Label	Bi-LSTM			BERT		
	Precision	Recall	F1 score	Precision	Recall	F1 score
Positive	0.86	0.86	0.86	0.96	0.93	0.94
Neutral	0.88	0.94	0.91	0.96	0.98	0.97
Negative	0.90	0.82	0.86	0.94	0.95	0.95
Weighted average	**0.876**	**0.878**	**0.877**	**0.952**	**0.952**	**0.952**

Enhanced Classification

Aspect Based Classification

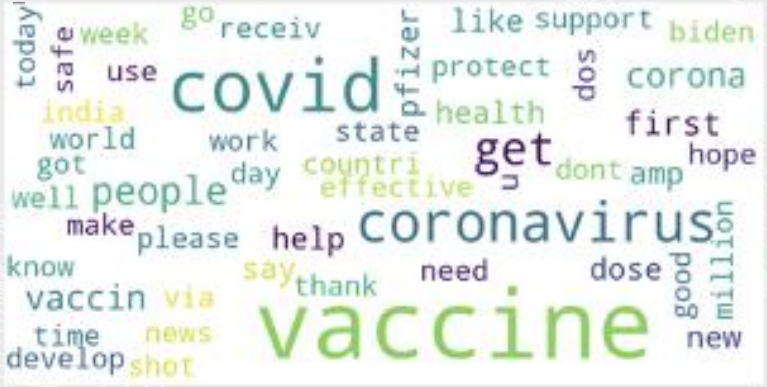
Model	Metrics		
	Precision	Recall	F1 score
Bi-LSTM	0.876	0.878	0.877
Attention Bi-LSTM	**0.903**	**0.903**	**0.903**

Ensemble Learning

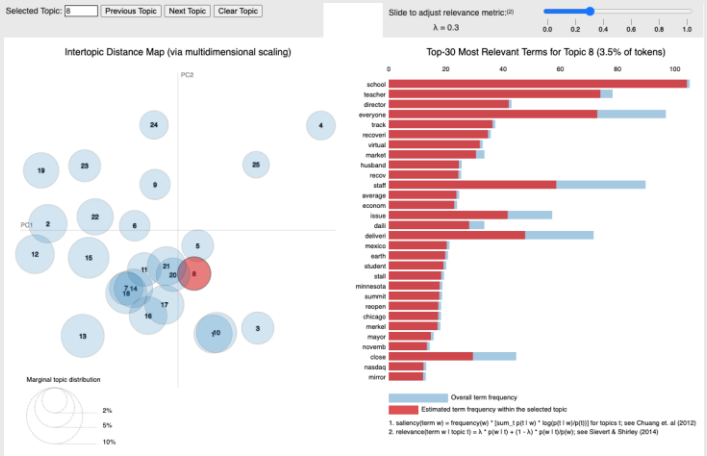
Model	Tf-idf Vectorizer		
	Precision	Recall	F1 score
Decision Tree	0.790	0.790	0.790
Random Forest	**0.905**	**0.904**	**0.903**

Visualisation

Word Cloud



LDA Topic Modelling



Geospatial Analysis

05.

CONCLUSION

Summary

Crawling

60,775 tweets crawled

Querying & Indexing

- Indexed tweets using Solr
- Web interface using Django

Classification

- Data Pre-processing
- Classified tweets into **Positive** or **Negative**
- Explored various models
- Enhanced classification with Ensemble Learning and Aspect Based Classification

THANKS!