



Tribhuvan University
Institute of Engineering
Pulchowk Campus



Social Media Mining for Health 2022 (#SMM4H)

COVID-19-related Nepali Tweets Classification in a Low Resource Setting

Rabin Adhikari, Safal Thapaliya, Nirajan Basnet, Samip Poudel,

Aman Shakya, Bishesh Khanal





COVID-19
RESPONSE

"COVID-19 is not just this century's largest public health emergency, but also a communication crisis."

UN Under-Secretary-General for Global Communications, Melissa Fleming
21 May 2020



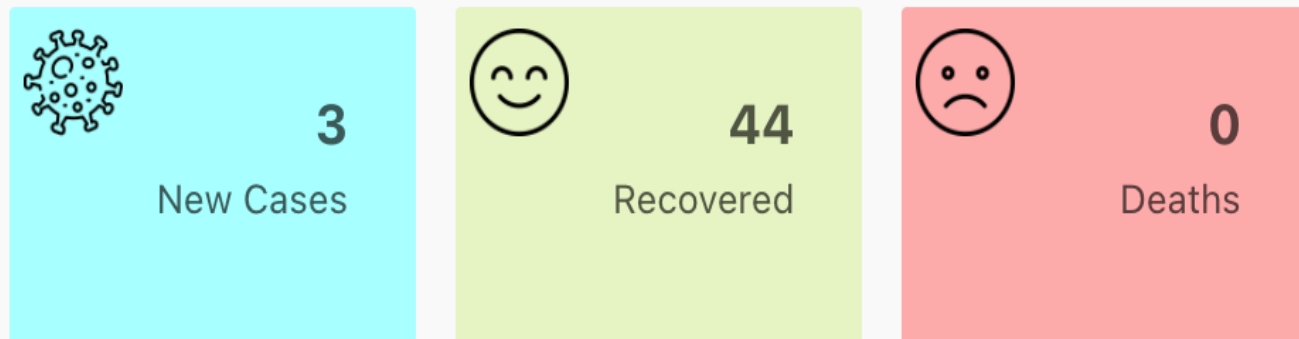
United
Nations

un.org/coronavirus | #COVID19

Nepal | Recent Updates

Last 24 hours update

Updated at: 14th April 2022



Download Situation Report

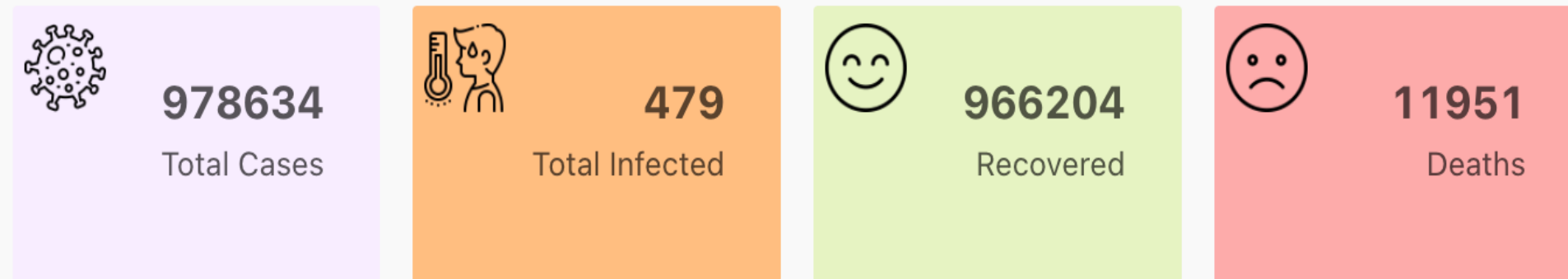


Take COVID-19 Self assessment test




Total Data

Updated at: 14th April 2022



 **5588749** | PCR Test

 **479** | Isolation

 **106** | Quarantined

Covid-19 Cases Map Representation

Filters

Select Province



Select Age Group



Select Gender



See Covid Hospitals/Lab

Start date



End date



Cases information 😊

HOSPITAL BED, ICU

HEOC

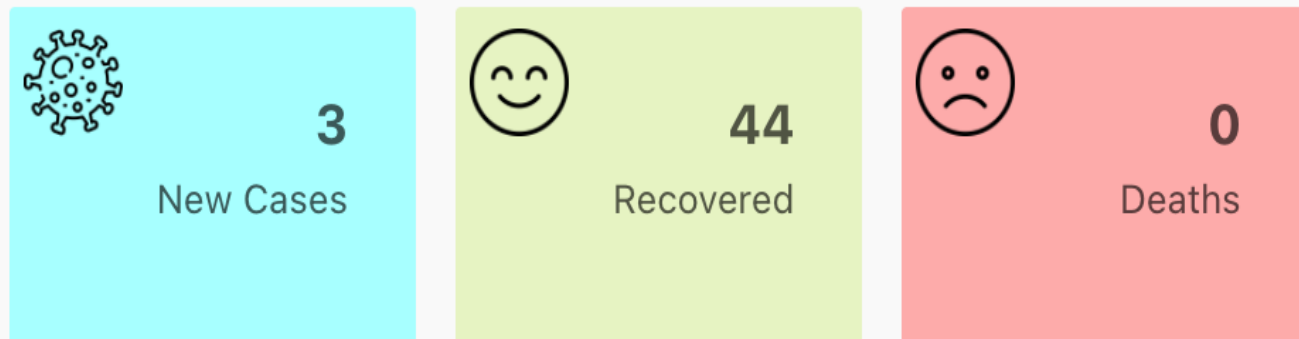
MOHP

Public Discourse?? 😞

Nepal | Recent Updates

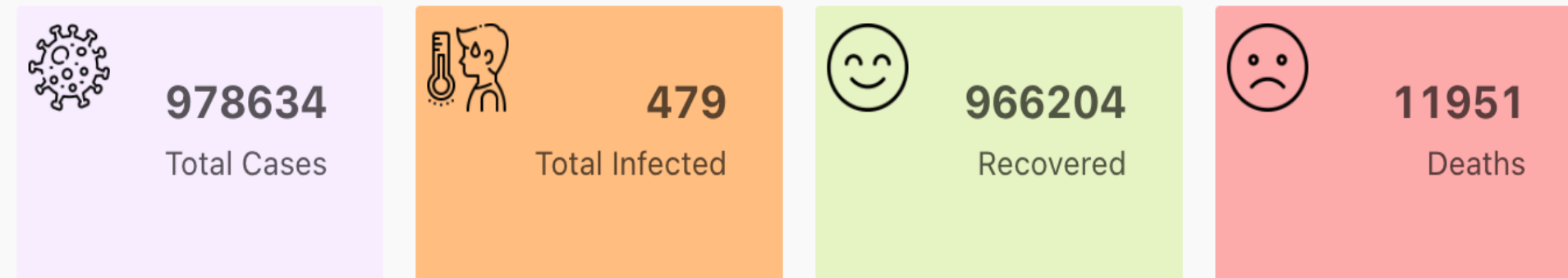
Last 24 hours update

Updated at: 14th April 2022



Total Data

Updated at: 14th April 2022



Download Situation Report




Take COVID-19 Self assessment test



 **5588749** | PCR Test

 **479** | Isolation

 **106** | Quarantined

Covid-19 Cases Map Representation

Filters

Select Province



Select Age Group



Select Gender



See Covid Hospitals/Lab

Start date



End date



SEE DATA FROM:

Last 7 days

All categories

All opinions



Early AI-supported Response with Social Listening

COVID-19 online conversations in 30 pilot countries

PURPOSE

Listening to people's questions and concerns is an important way for health authorities to learn about what matters to communities in response to COVID-19. This social listening platform aims to show real time information about how people are talking about COVID-19 online, so we can better manage as the infodemic and pandemic evolve.

[Learn more about the infodemic](#)

DATA

 **75,777,362**

posts analyzed since 15 December 2020

 **30**

countries covered

 **9**

languages analyzed

 **41**

COVID-19 categories tracked

[More about the data, limitations and how to interpret it](#)

What are people talking about?

[Overview across countries](#)[Trend of intensity per](#) [country](#) [region](#)

SEE DATA FROM:

Last 7 days

All categories

All opinions



Early AI-supported Response with Social Listening

COVID-19 online conversations in 30 pilot countries

Public Discourse in Nepal?? 😞

PURPOSE

Listening to people's questions and concerns is an important way for health authorities to learn about what matters to communities in response to COVID-19. This social listening platform aims to show real time information about how people are talking about COVID-19 online, so we can better manage as the infodemic and pandemic evolve.

[Learn more about the infodemic](#)

DATA

 **75,777,362**

posts analyzed since 15 December 2020

 **30**

countries covered

 **9**

languages analyzed

 **41**

COVID-19 categories tracked

[More about the data, limitations and how to interpret it](#)

What are people talking about?

[Overview across countries](#)[Trend of intensity per](#) [country](#) [region](#)

Proposed Solution

Web-based Application

- Display and visualize the tweets labelled by humans and the Machine Learning model.
- Show the discourse using visualizations.



Outline



Problem Statement



Objectives



Implementation Details



Results



Future Enhancements

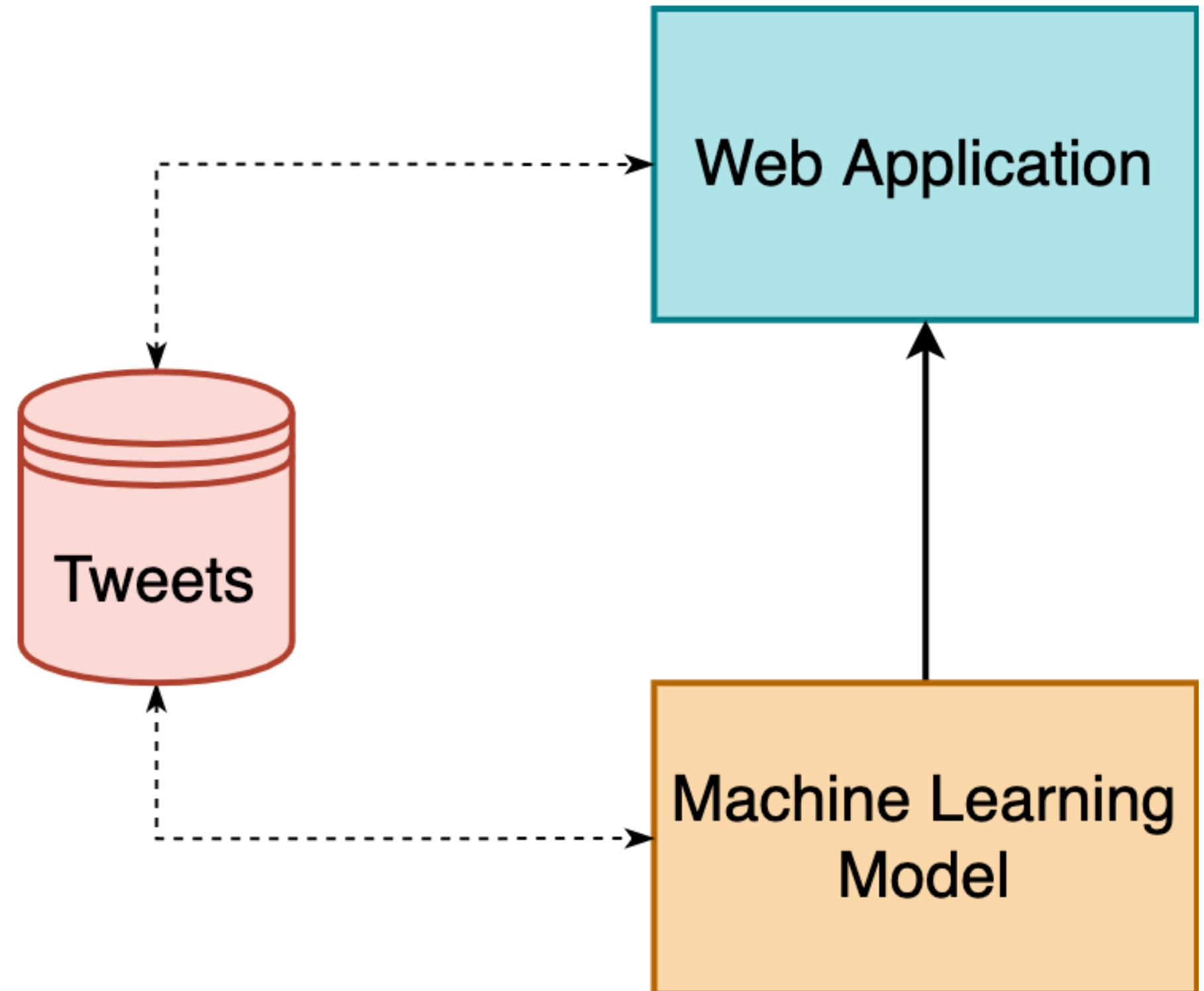
Project Composition

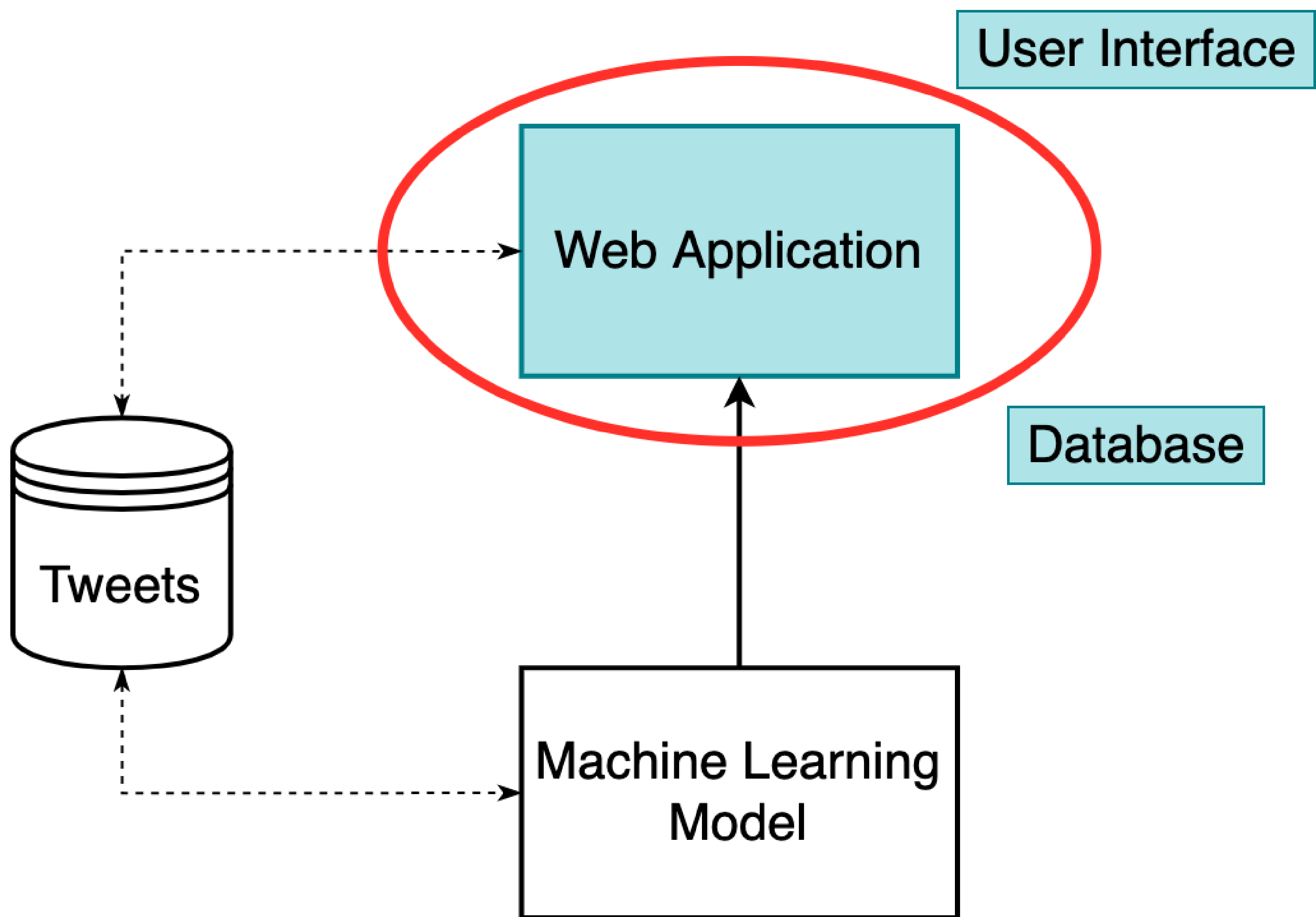
Web Application

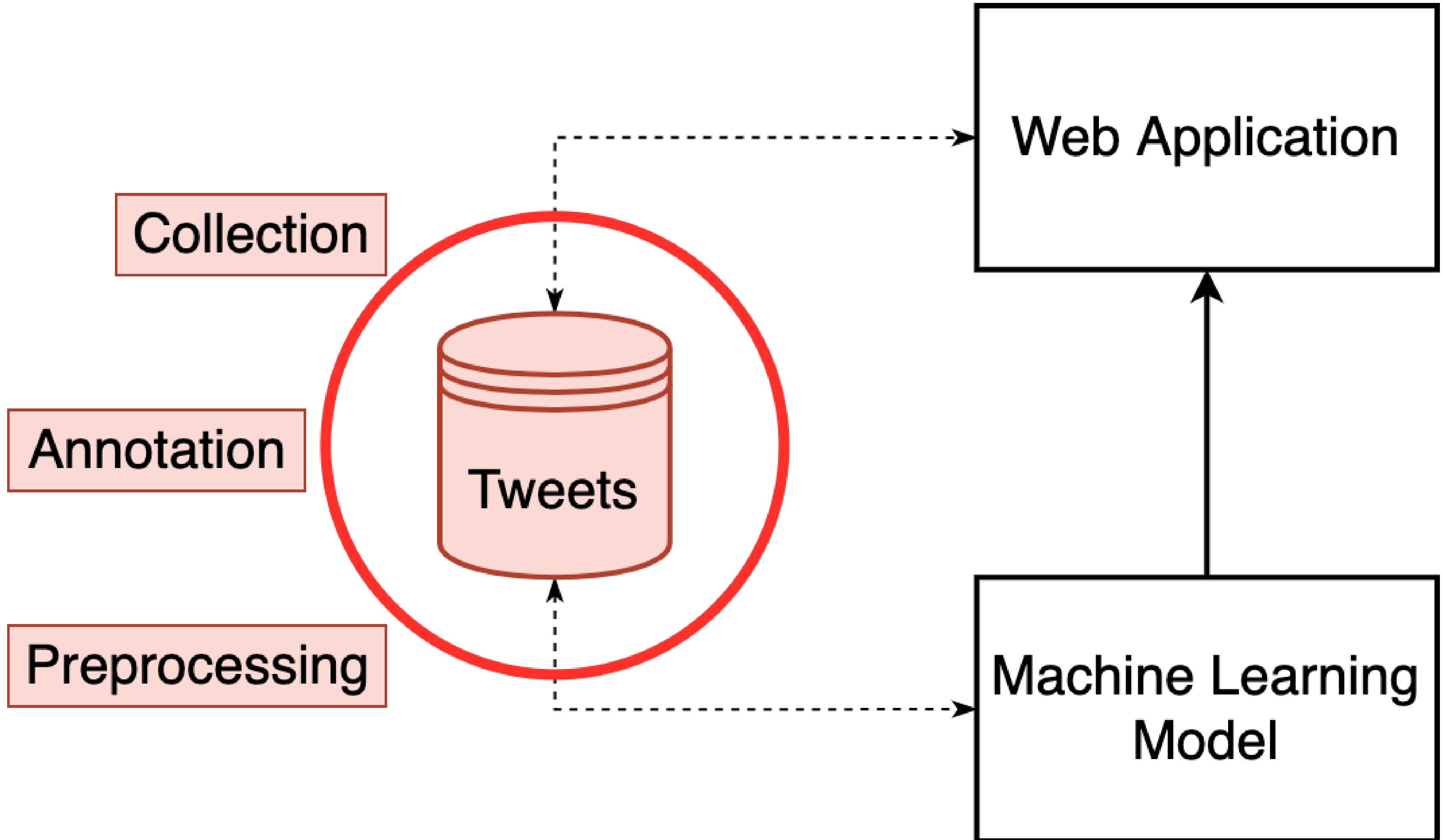
Display and visualize the tweets labelled by human and the Machine Learning model.

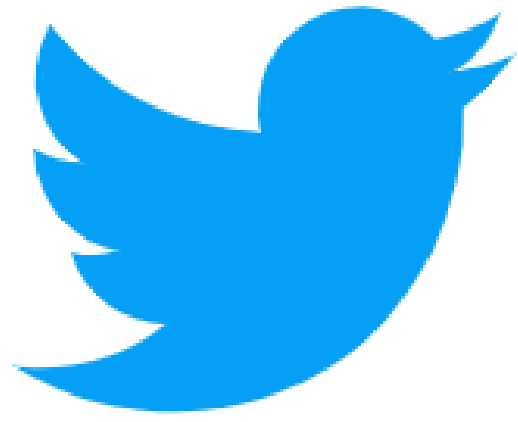
Machine Learning Model

Provide predictions for the topics of individual tweets.

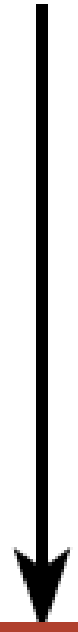








Nepali Language Tweets



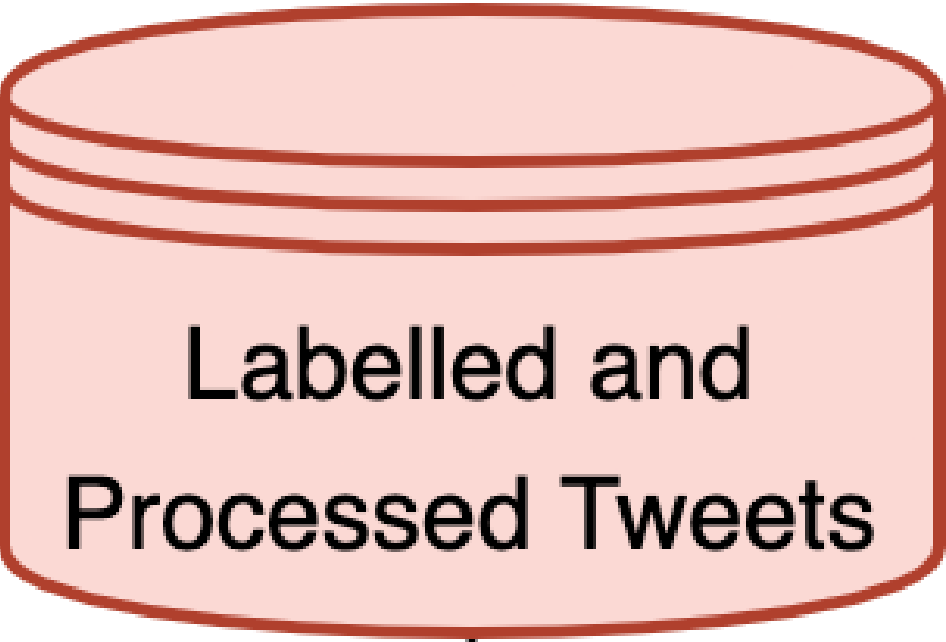
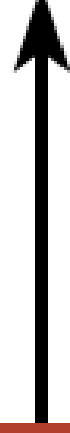
**Tweets collection
using keywords**



**Tweet
Annotation**



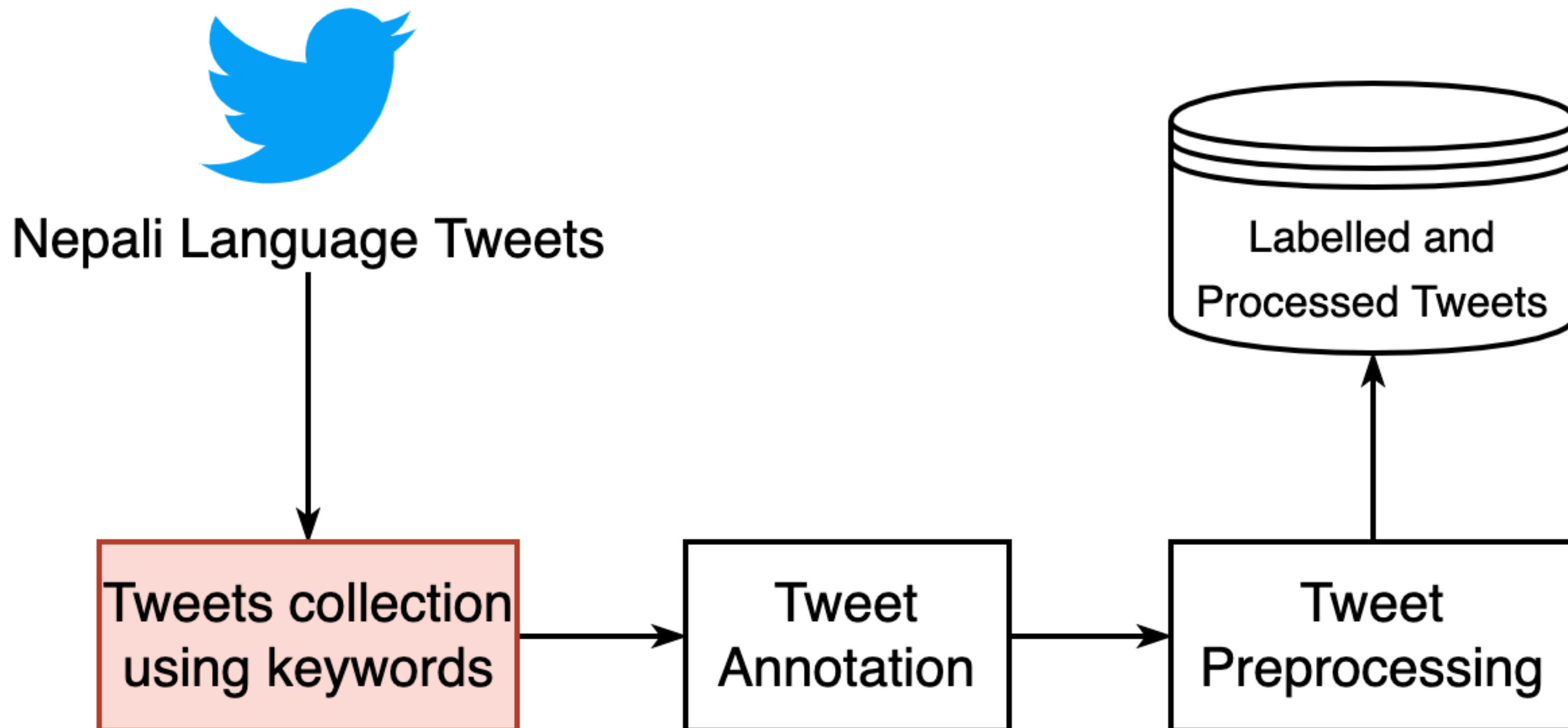
**Tweet
Preprocessing**



**Labelled and
Processed Tweets**

Keywords

कोभिड, कोरोना, कोरोनाभाइरस, कोरोनाभाईरस, कोवीड, कोभीड, कोविड, भ्याक्सिन, भयाक्सीन, बेड, "अक्सिजन सिलिन्डर", लकडाउन, निशेधाज्ञा, संक्रमण, संक्रमित, खोप, फाइजर, फाईजर, भेरोसेल, भेरोशेल, "जोर बिजोर", जोर-बिजोर, "स्वास्थ्य तथा जनसंख्या मन्त्रालय", "होम आइसोलेसन", आइसोलेसन, आईसोलेसन, "दोस्रो लहर", प्रभावकारी, प्रभावकारिता, महामारी, माहामारी, भेरियन्ट, भेरीयन्ट, डेल्टा, "डेल्टा प्लस", "संक्रमण मुक्त", मास्क, स्यानिटाइजर, "भौतिक दुरी", डोज, म्युटेसन, शय्या, जोखिम, आइसियु, भेन्टिलेटर, एन्टिजेन, कोभ्याक्स, "विश्व स्वास्थ्य संगठन"



- 8 topics derived from WHO EARS
- 7 annotators



Nepali Language Tweets

Tweets collection
using keywords

Tweet
Annotation

Tweet
Preprocessing

Labelled and
Processed Tweets

Topics

- Covid Stats
- Vaccination
- Covid Politics
- Humor
- Lockdown
- Civic Views
- Life During Pandemic
- Waves and Variants

कोरोना संक्रमणको दर बढ्यो ~~<https://t.co/zx0inwXmpr> via @himal_dainik~~



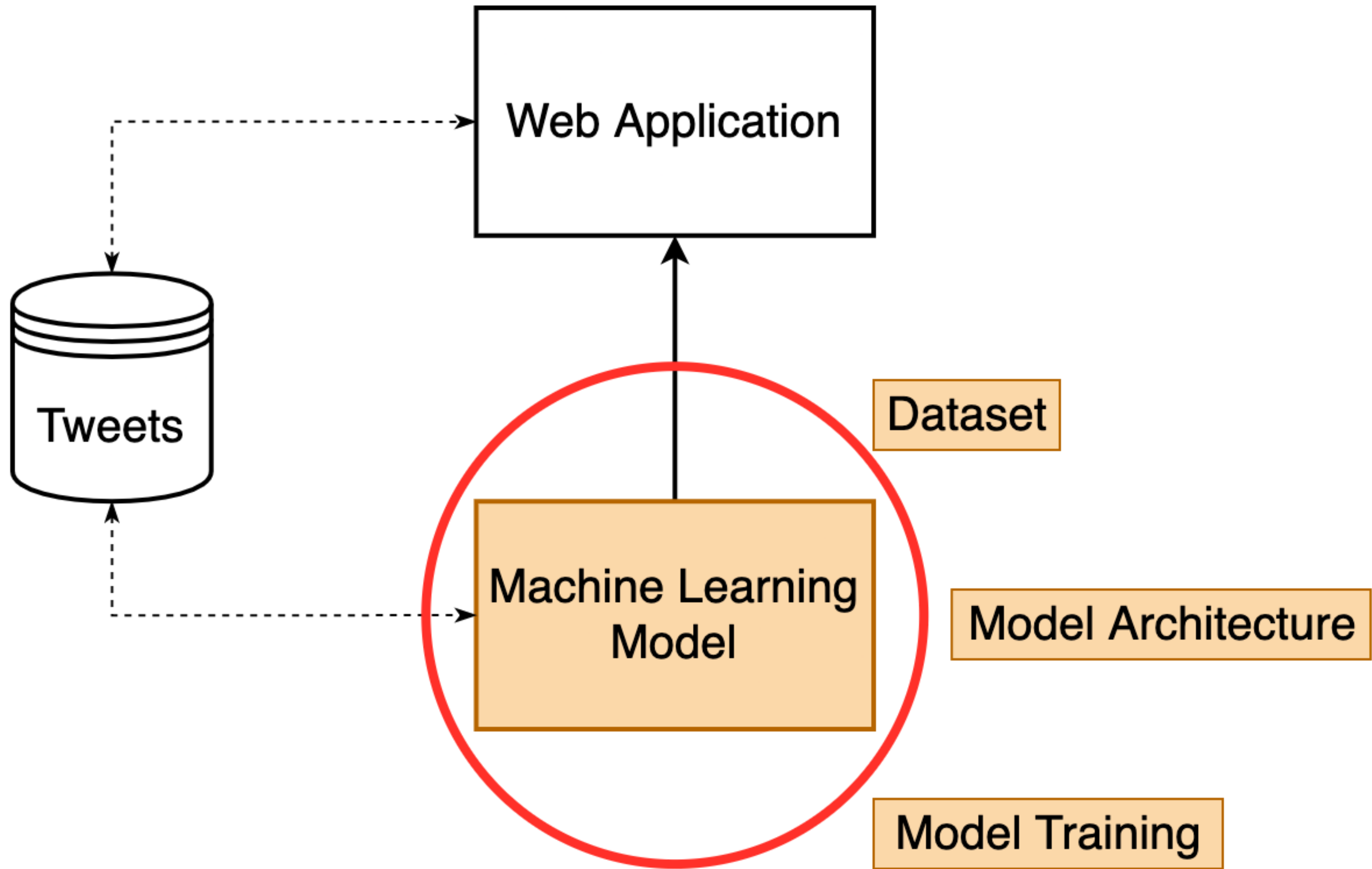
Nepali Language Tweets

Tweets collection
using keywords

Tweet
Annotation

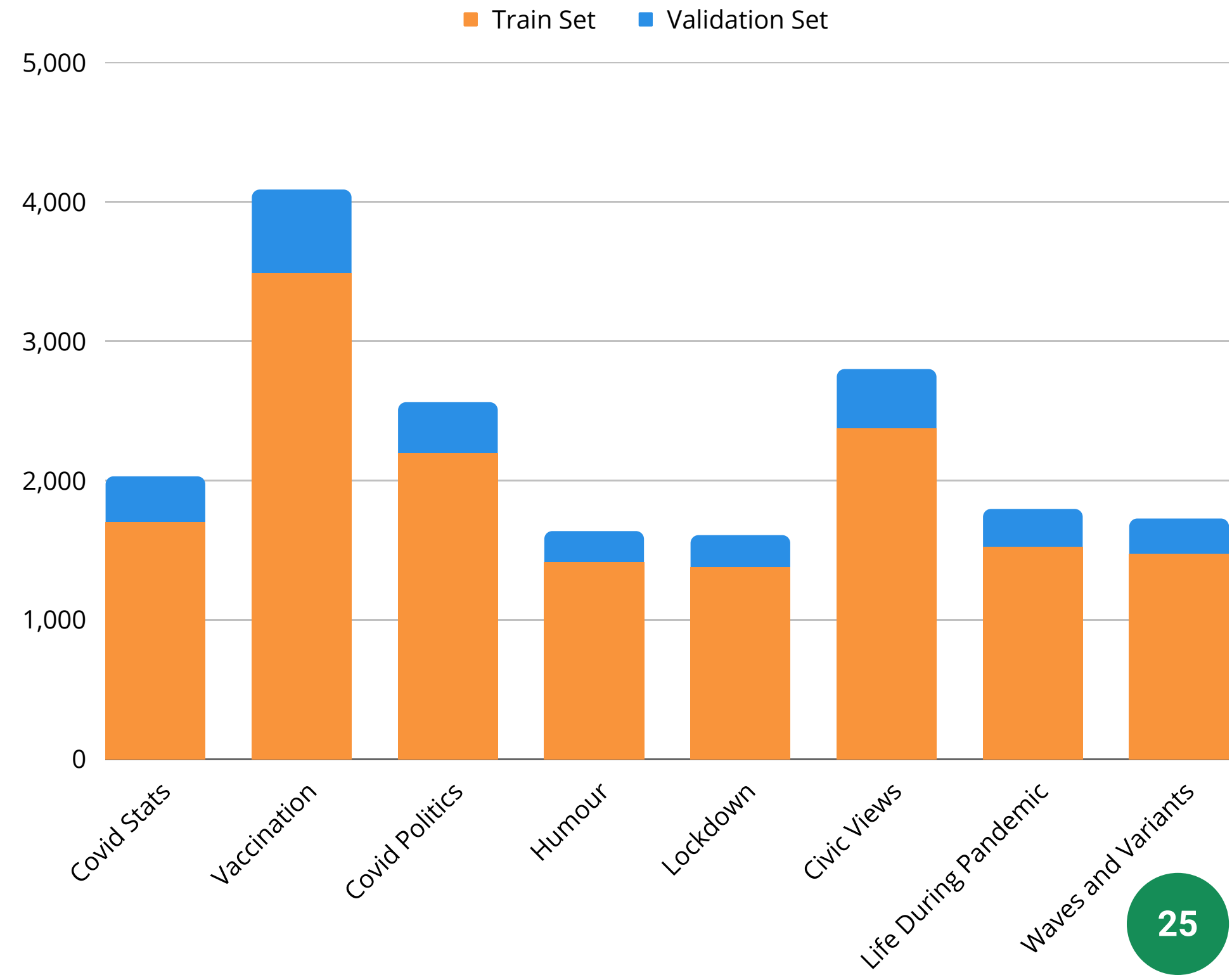
Tweet
Preprocessing

Labelled and
Processed Tweets

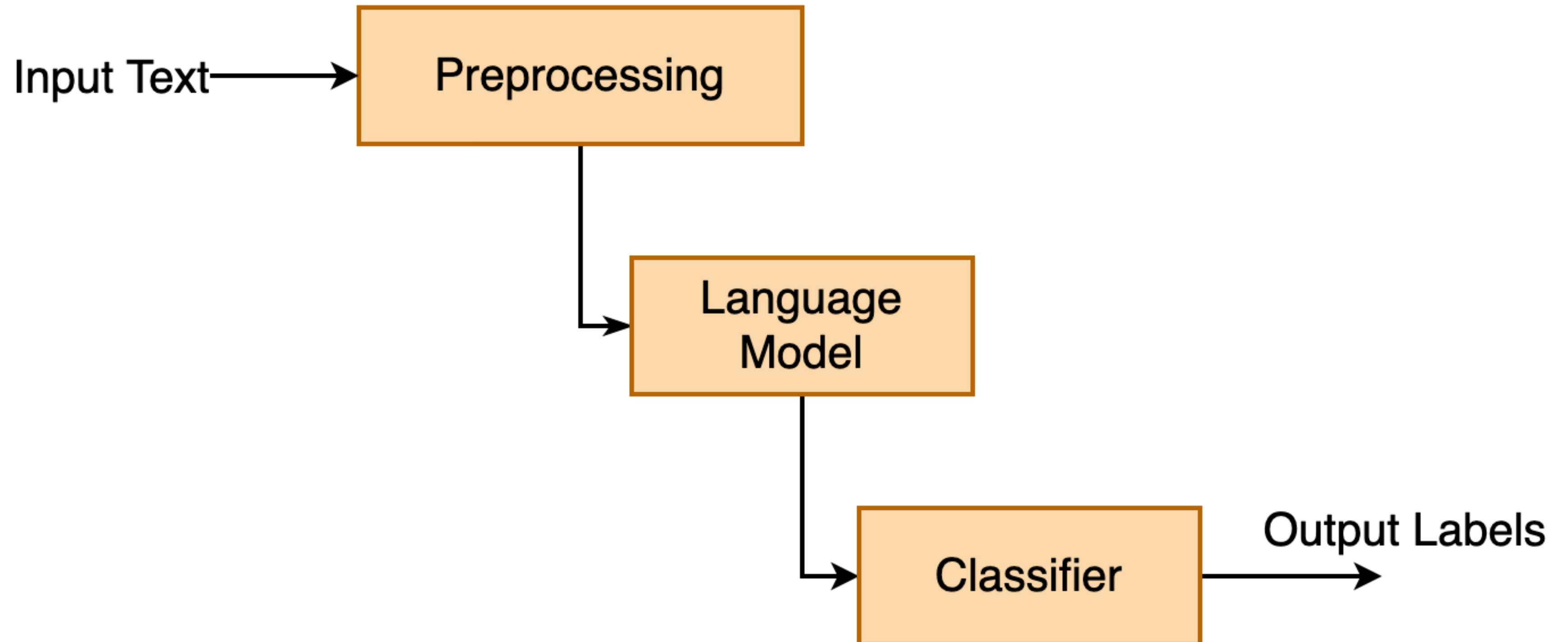


NAT-CTC Dataset

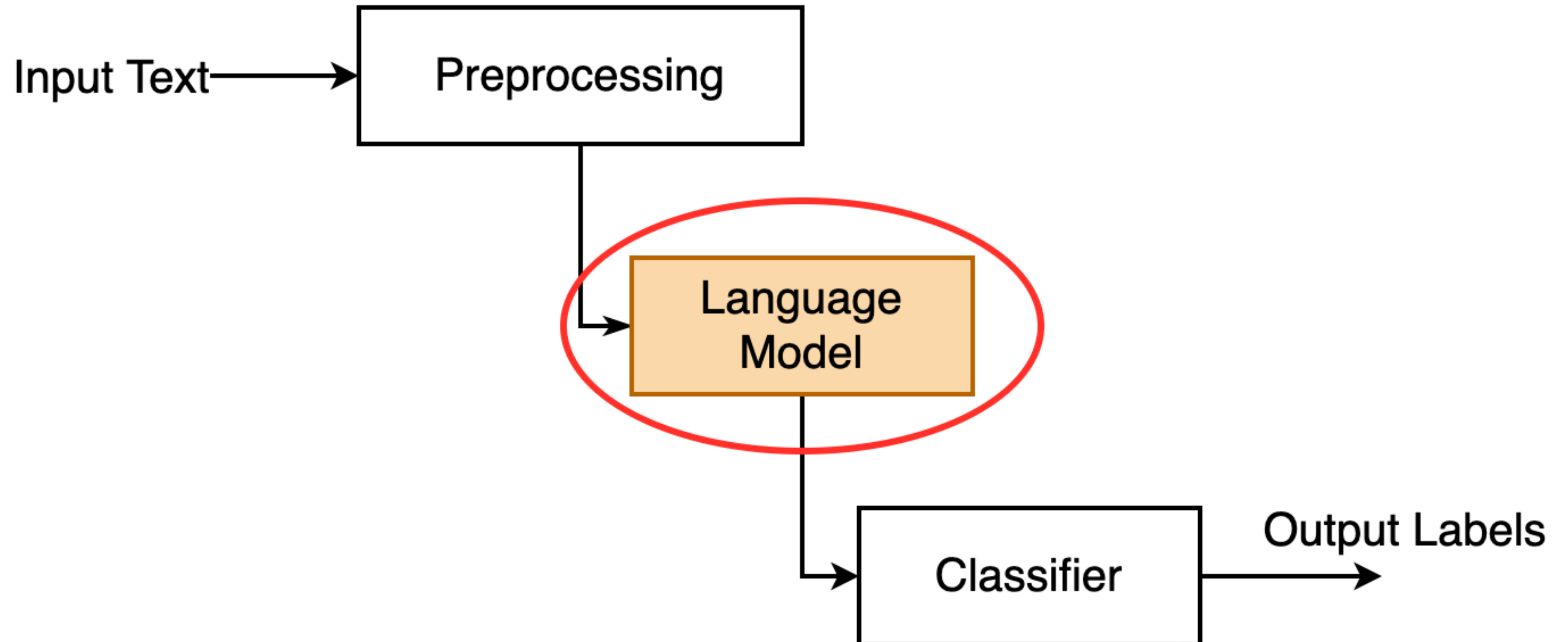
- A total of 12,241 labelled tweets
- 15% Validation Set
- Topics Inter-rater agreement
 - Fleiss Kappa score of 0.64



Machine Learning Model: Architecture



Machine Learning Model: Architecture



Language Model Choices

- Poor Performance of Nepali Language Models
- Choice narrowed down to **Multilingual Models**
- **mBERT**: More Latin Scripts ~104 languages
- **MuRIL**: More emphasis on Indic Languages ~ 17 languages

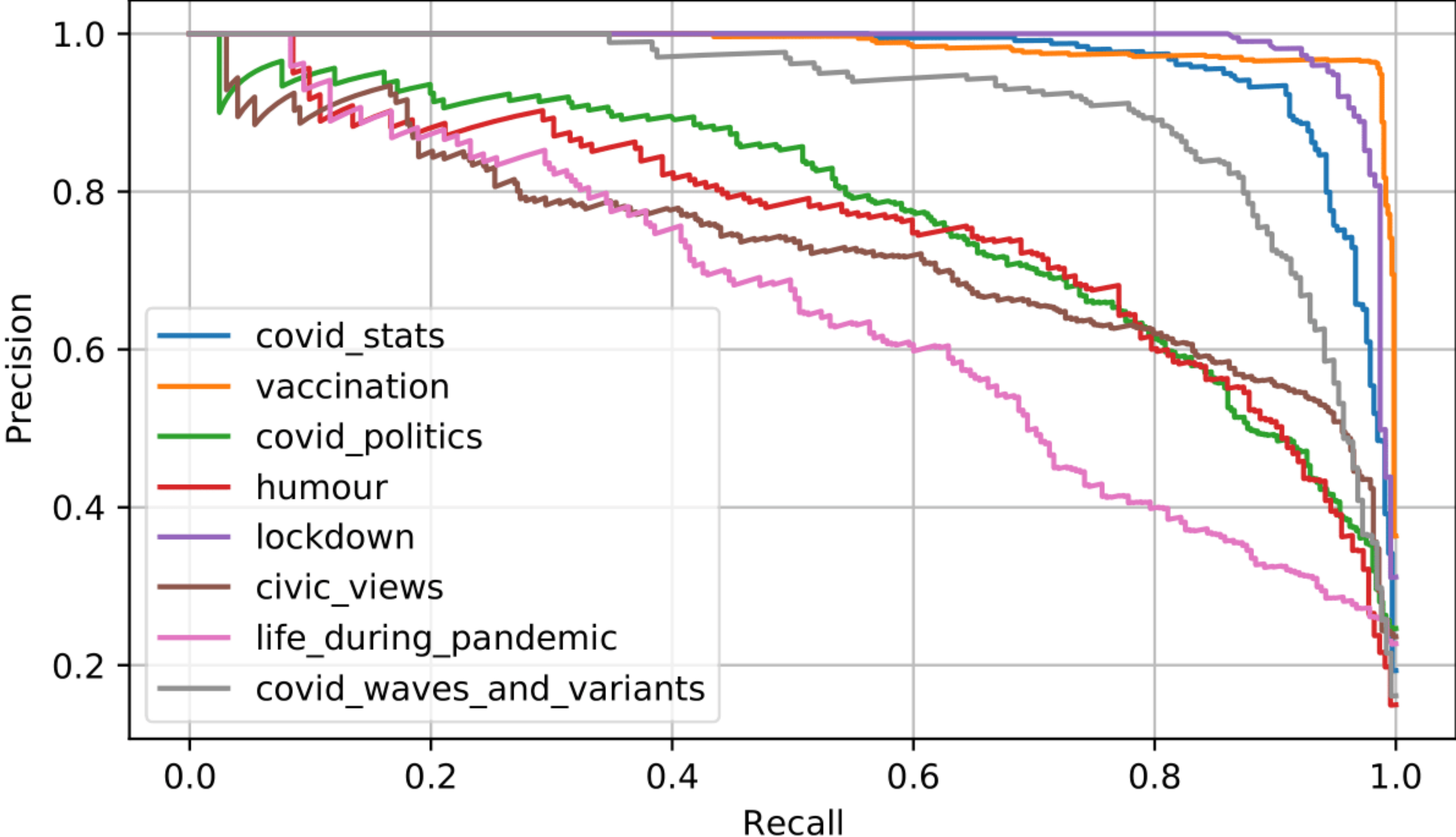
Devlin et. al., *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, 2018

Khanuja et. al. *MuRIL: Multilingual Representations for Indian Languages*, 2021

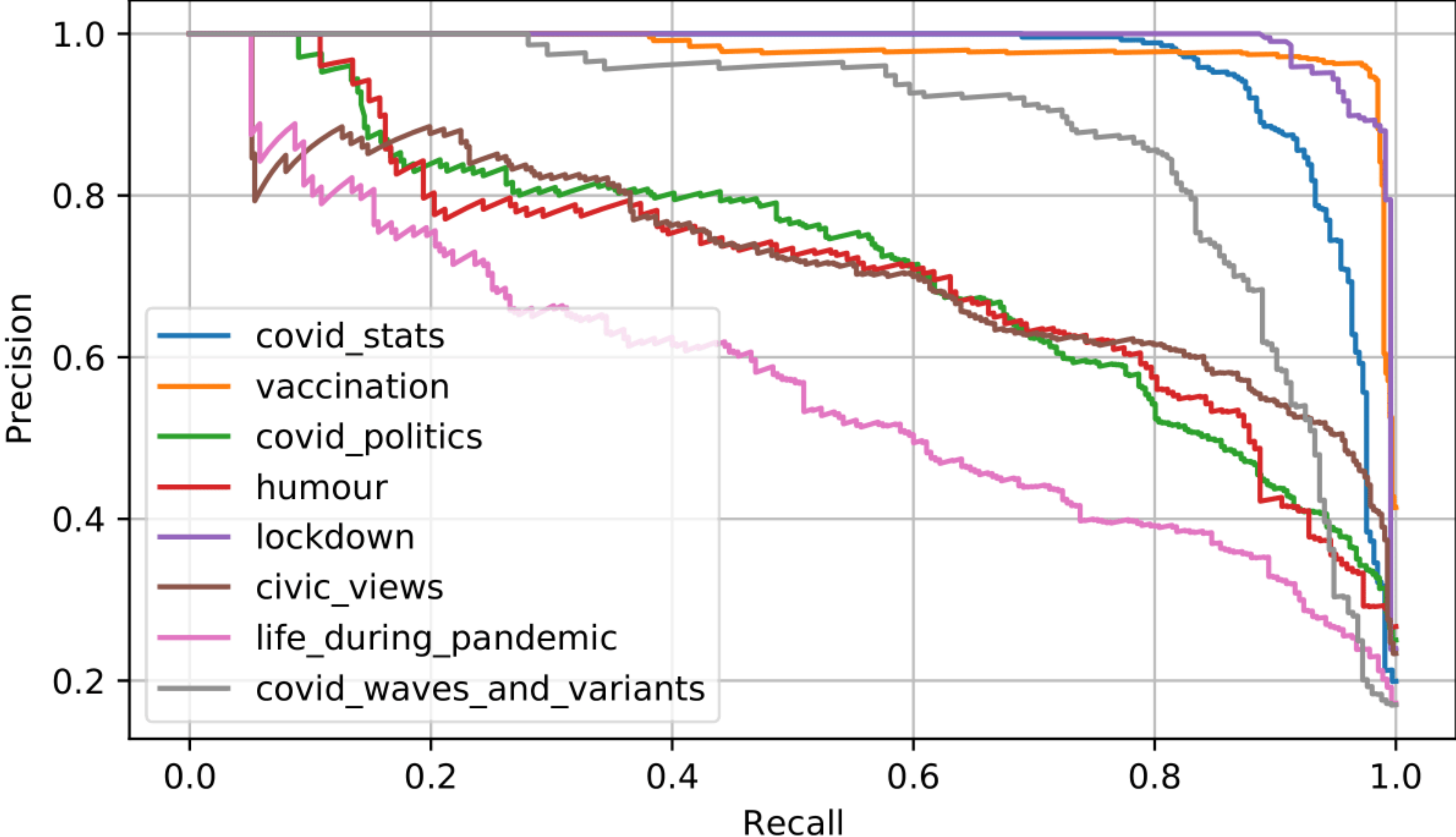
Outline

- Problem Statement
- Objectives
- Implementation Details
- 4** ● **Results**
- Future Enhancements

Precision-Recall Curve: MuRIL



Precision-Recall Curve: mBERT



MuRIL vs mBERT

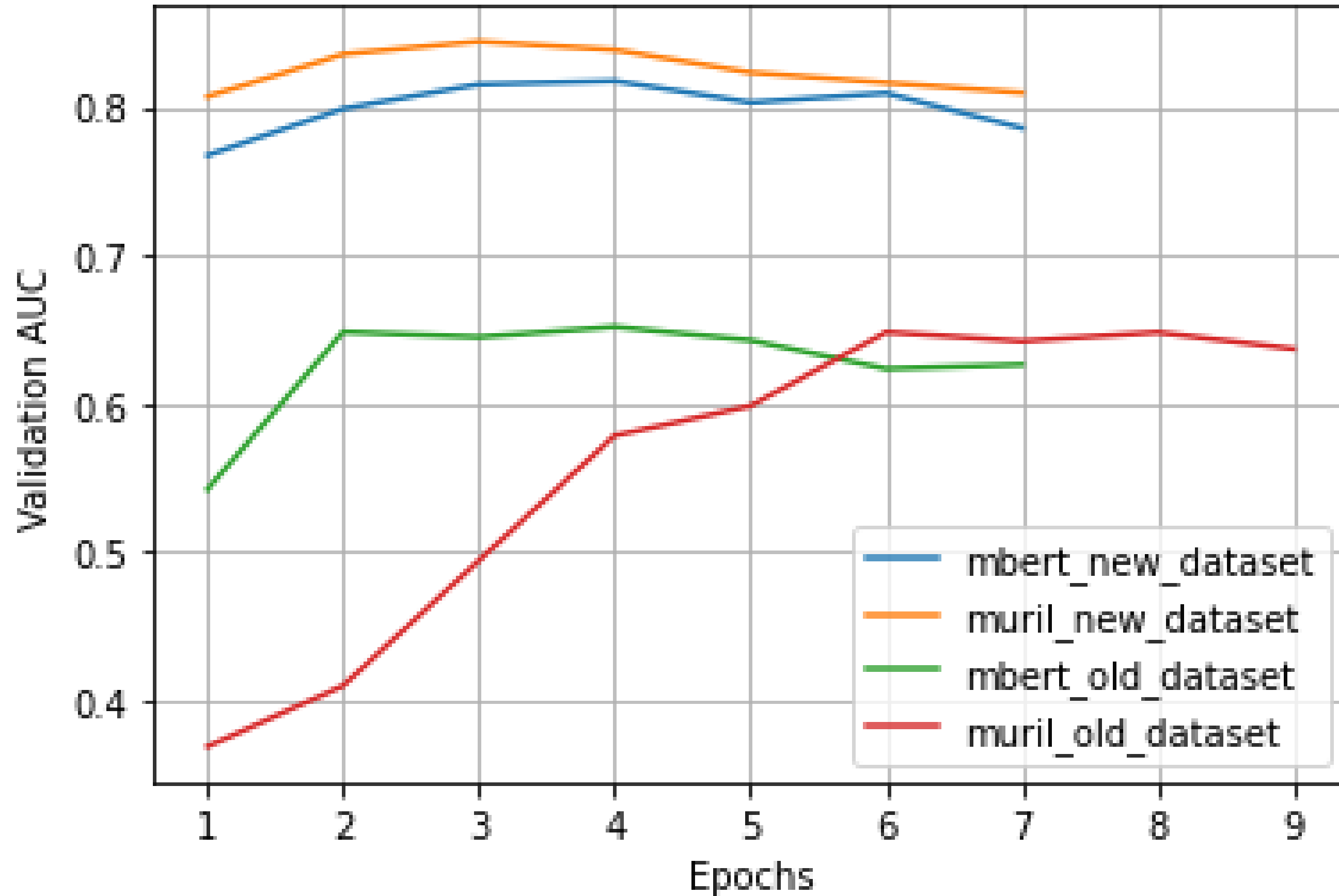
AUC

Labels	MuRIL	mBERT
COVID Stats	0.97	0.96
Vaccination	0.99	0.98
COVID Politics	0.77	0.72
Humor	0.756	0.71
Lockdown	0.986	0.989
Civic Views	0.73	0.72
Life During Pandemic	0.65	0.57
COVID Waves and Variants	0.91	0.88

F1-Score

MuRIL	mBERT
0.92	0.91
0.974	0.968
0.71	0.67
0.72	0.69
0.952	0.95
0.702	0.701
0.62	0.55
0.85	0.83

MuRIL > mBERT on Bigger Dataset



Outline

- Problem Statement
- Objectives
- Implementation Details
- Results
- 5** ● **Future Enhancements**

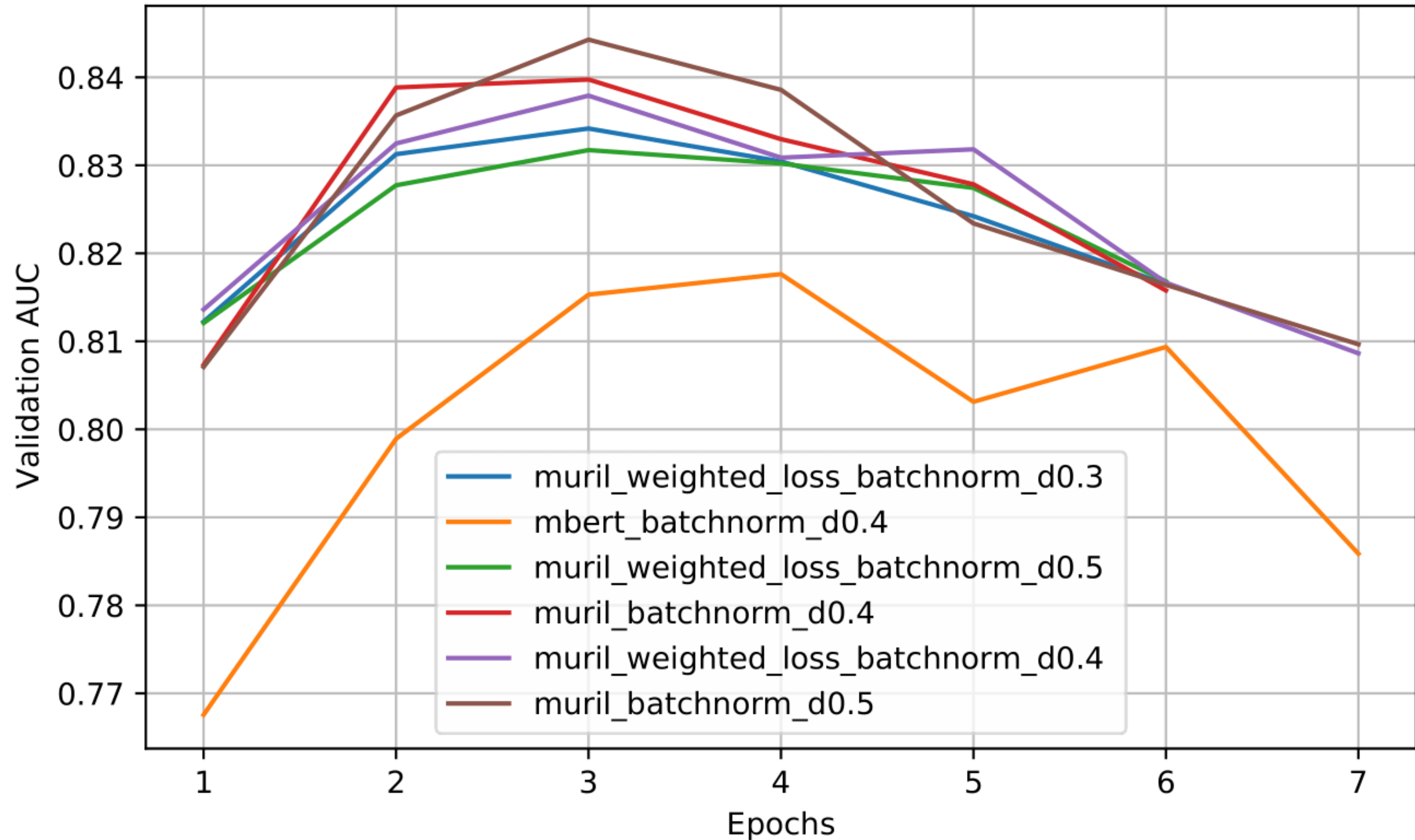
Future Works

- Analyze the basic sentiments of the people during the epidemic
- Add hierarchical categories
- Determine misinformation
- Recommendation system for doctors and experts

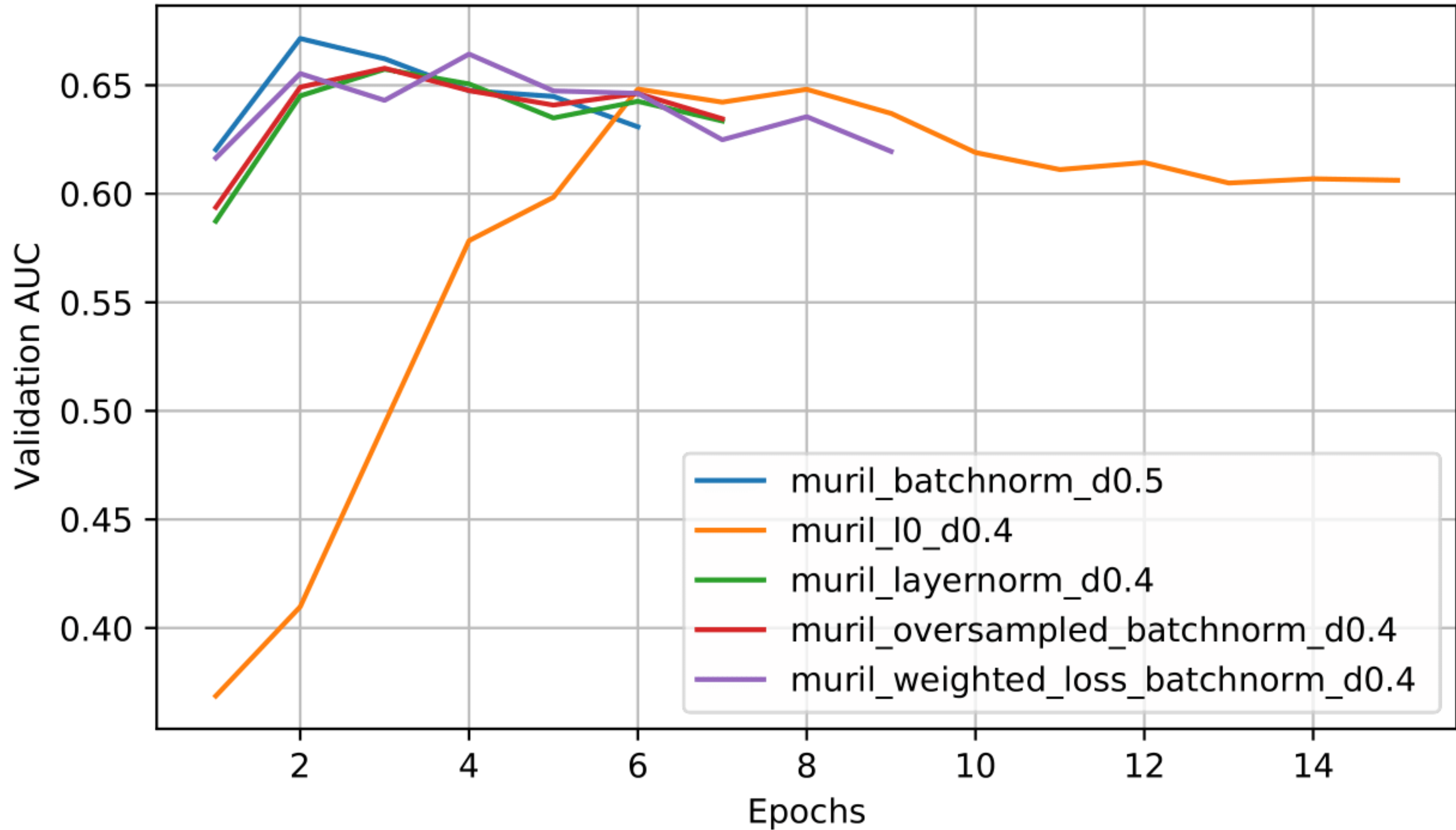
The background of the image is a green-tinted photograph of a group of hikers on a mountain trail. The hikers are silhouetted against a bright sunset or sunrise, with a glowing sun visible in the upper left. The hikers are wearing backpacks and using trekking poles. Some are pointing towards the horizon, and one in the foreground is looking back over their shoulder. The overall mood is one of accomplishment and gratitude.

Thank You

Models on Bigger Dataset



Models on Smaller Dataset



Best Model Architecture

