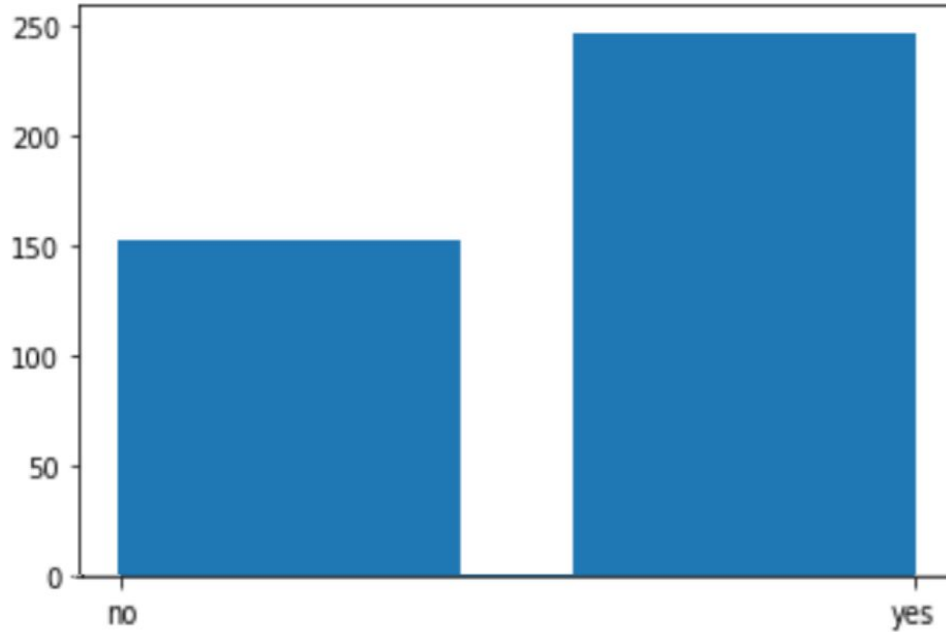


Covid-19 Fact Checking

by

Khang Zheng Ng, Michael Naccache
and Amina Isse

Training Set Distribution



Distribution		
Class	Amount	Percentage
no	152	38.1
yes	247	61.9
total	399	100

Model Differences

Original Vocabulary	Filtered Vocabulary
“One model where all words appearing in the training set are used as features.”	“A second model where you filter out the words that appear only once in the training set, so V contains only the words that appear at least 2 times in the training set.”
<code>count_vect = CountVectorizer()</code>	<code>count_vect = CountVectorizer(min_df=2)</code>

Part 1: Original Vocabulary

Confusion Matrix and Classification Report

	precision	recall	f1-score	support
no	0.79	0.52	0.63	21
yes	0.75	0.91	0.82	33
accuracy			0.76	54
macro avg	0.77	0.72	0.73	54
weighted avg	0.76	0.76	0.75	54

```
[[11 10]
 [ 3 30]]
```

Part 2: Filtered Vocabulary

Confusion Matrix and Classification Report

	precision	recall	f1-score	support
no	0.80	0.57	0.67	21
yes	0.77	0.91	0.83	33
accuracy			0.78	54
macro avg	0.78	0.74	0.75	54
weighted avg	0.78	0.78	0.77	54

```
[[12  9]
 [ 3 30]]
```

Part 1: LSTM Run 1

Evaluation Metrics

Accuracy_score: 0.7037

precision_score yes: 0.7179 precision_score no: 0.6667

recall_score yes: 0.8485 recall_score no: 0.4762

f1_score yes: 0.7778 f1_score no: 0.5556

Part 2: LSTM Run 2

Evaluation Metrics

Accuracy_score: 0.7407

precision_score yes: 0.7111 precision_score no: 0.8889

recall_score yes: 0.9697 recall_score no: 0.381

f1_score yes: 0.8205 f1_score no: 0.5333