Fake News Detection

By SA3



TABLE OF CONTENTS

01

BACKGROUND

03

MODEL

02

DATASET

04

RESULT





ISSUE AND EXISTING MEASURES

Fake news is causing undesirable public alarm!

Current Measures:

Educating people about fake news

Will people actually know if a news is fake? What if they didn't get the chance to learn about fake news identification?

Identifying Fake News

About half of the people polled here - 48 to 53 per cent - said they could tell if a piece of information on social media is true or false.

However, about seven in 10 - 69 to 76 percent - admitted that they have unknowingly shared fake news, according to the study by the Nanyang Technological University's (NTU) Centre for Information Integrity and the Internet (IN-cube).

This is dangerous because confident people might end up believing information they come across that they do not have the ability to figure out is true or not, said Associate Professor Edson Tandoc Jr, director of IN-cube.







Dataset



COVID-19 Fake News

"A video shows an Italian doctor who says that nobody died from coronavirus in Italy that it's not more dangerous than a flu."



US Politics Fake News

"If there's one thing that conservatives and Fox News want more than anything, it's for President Obama and his administration to fail ..."

Data Cleaning

- Remove html tags (<head>, </head>)
- 2. Remove text between square brackets ([0932 EST])
- 3. Remove stopwords ("the", "a", "an", "in")

Most common words



```
{'cases': 1427,
'COVID-19': 1191,
'#COVID19': 1151,
'new': 734,
'people': 642,
'coronavirus': 640,
'tests': 569,
'number': 516,
'confirmed': 471,
'reported': 449}
```

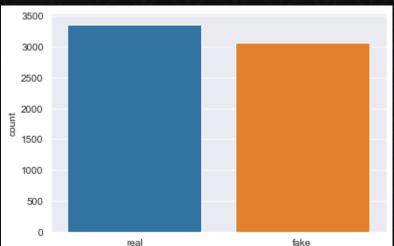


{'Trump': 111503, 'said': 93162, 'would': 54613, 'U.S.': 50441, 'President': 33180, 'people': 33115, 'also': 30325, 'one': 29370, 'Donald': 27795, 'said.': 26194}

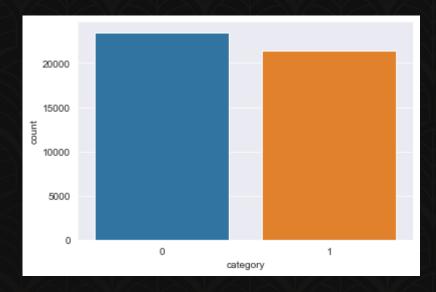
Are our datasets balanced?



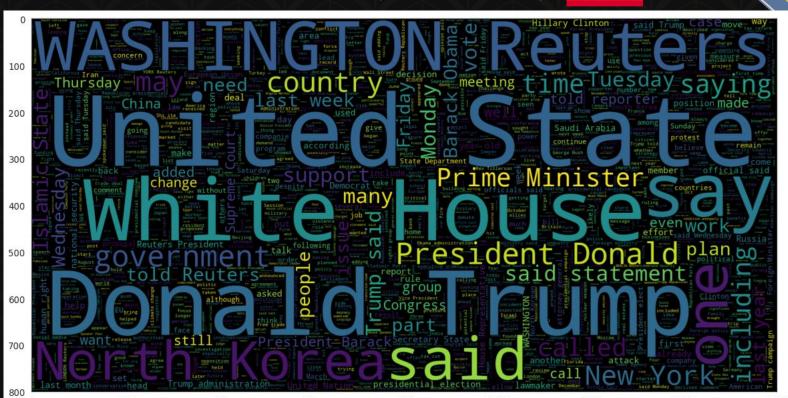




label



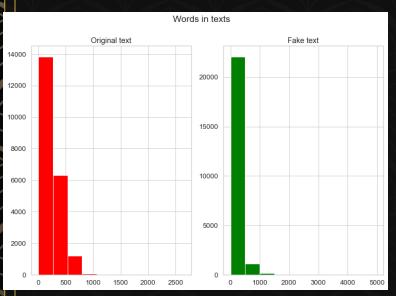
US Politics Fake News (Real Text)

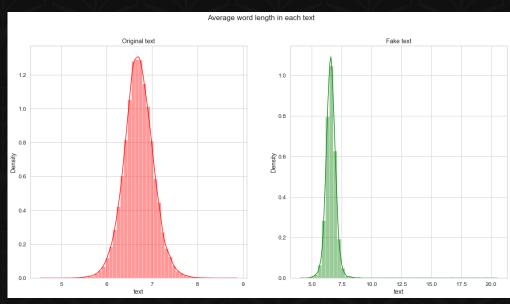


US Politics Fake News (Fake Text)

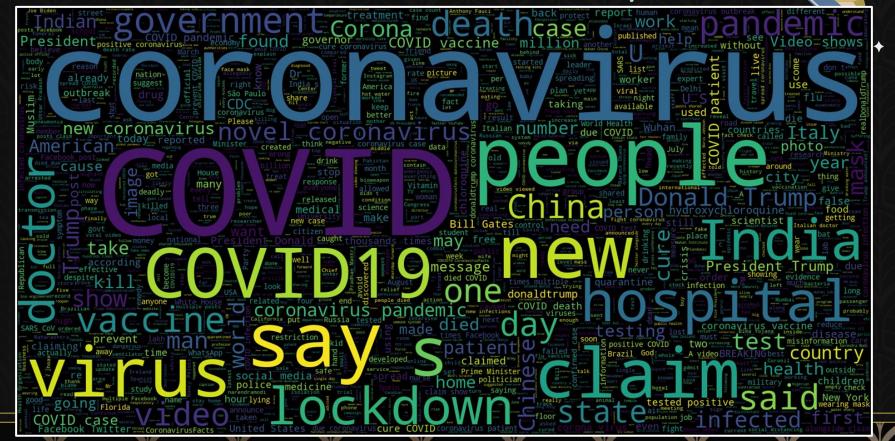


US Politics Number of Words/Word Length

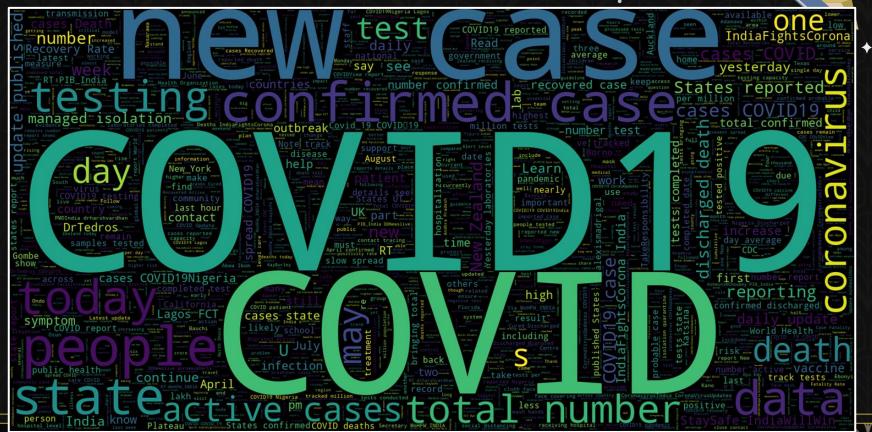




COVID-19 Fake News (Fake Text)

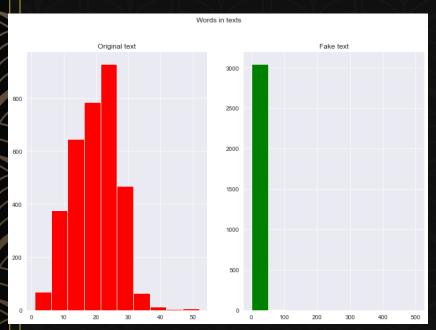


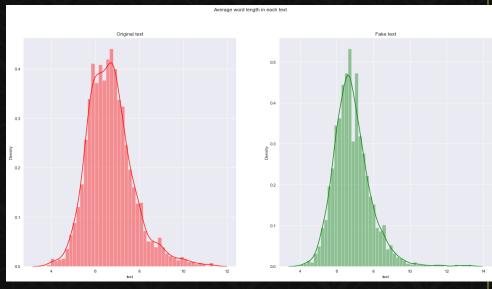
COVID-19 Fake News (True Text)



COVID-19 Number of Words/Word Length

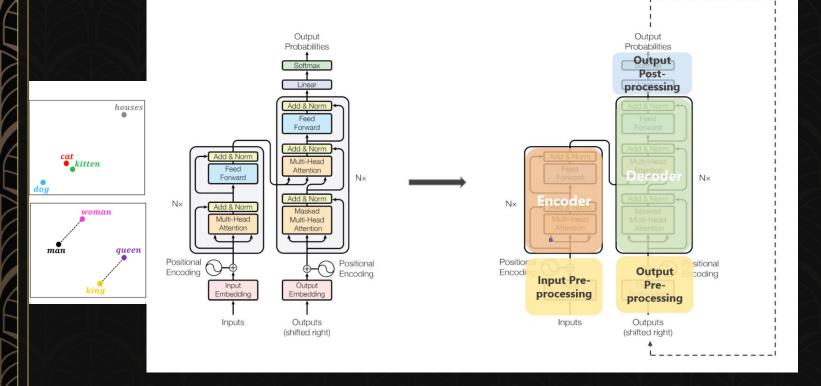




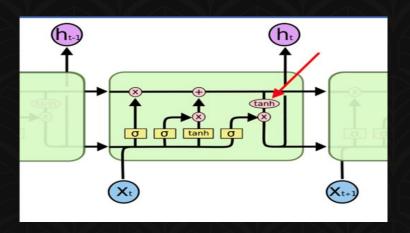




1. Transformer Model



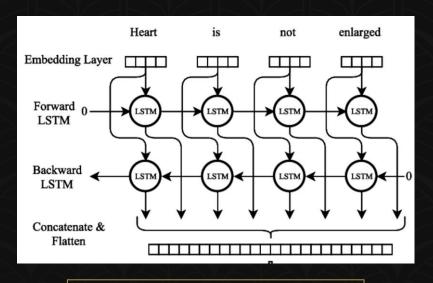
LSTM



Long-Short Term Model

Based on Recurrent Neural Network, with additions of cell gates that store or forgo memories

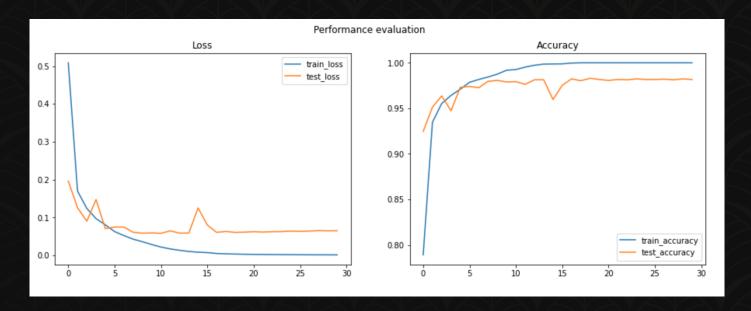
Bi-Directional LSTM



Long-Short Term Model

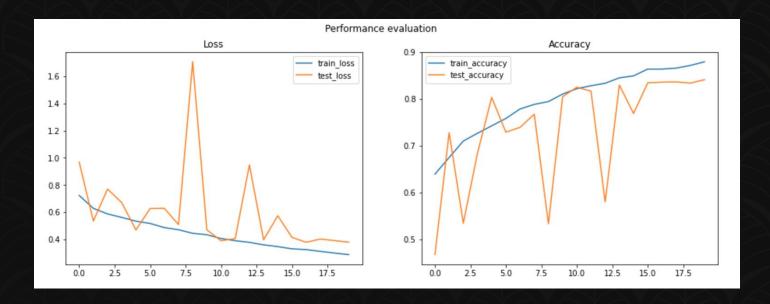


#1 Transformer - US Politics



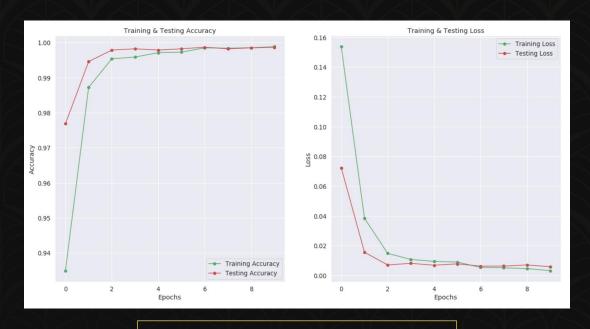
Accuracy: 98.2%

#1 Transformer - Covid-19



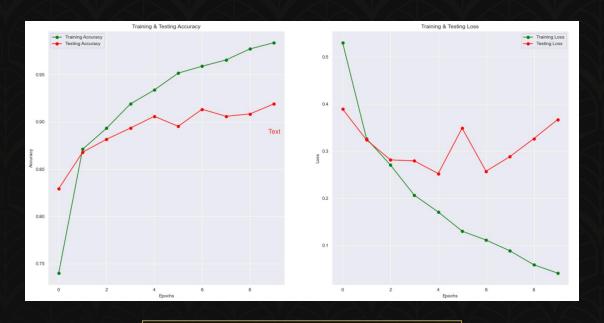
Accuracy: 84.1%

#2 LSTM - US Politics



Accuracy: 99.2%

#2 LSTM - Covid-19



Accuracy: 91.9%

#3 Bi-Directional LSTM Accuracy: 99.3% **US Politics** Accuracy: 97.6% Covid-19

Collation of Results

MODEL	Accuracy (US Politics)	Accuracy (Covid -19)
Transformer	98.2%	84.1%
LSTM	99.2%	91.9%
Bi-Directional LSTM	99.3%	97.6%

Collation of Results

MODEL	Accuracy (US Politics)	Accuracy (Covid -19)
Transformer	98.2	84.1
LSTM	99.2	91.9
Bi-Directional LSTM	99.3	97.6

Possible Use Cases

- Filter out fake social media news stories
- App to warn user about potential fake news
- Blacklist sources that have many fake news publications
- Help POFMA enforcement



