

A close-up, slightly blurred photograph of a stack of newspapers. The top newspaper's masthead 'WORLD BUSINESS' is clearly visible, along with the section code 'B3'. Below the masthead, some text from an article is partially legible, including '...countries have' and '...Gen. Tha'. The background is a soft, out-of-focus blue-grey.

Detecting Fake News on Social Media

Regina Soh

IOD Capstone Project

Agenda

1

Fake News

Problems &
Stakeholders

2

Exploratory Data Analysis

Insights from dataset

3

Modelling Results

Accuracy scores of
classification models

4

Conclusion

Key takeaway
& future work

Bio

- Education

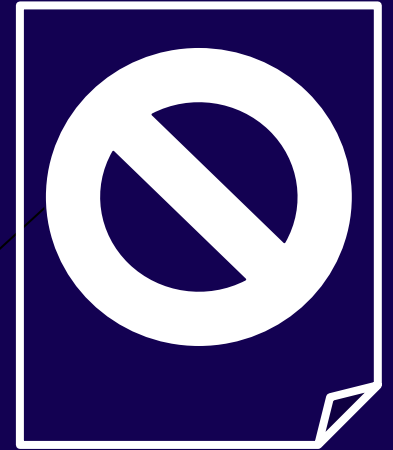
- Bachelor of Science in Life Sciences
- Master of Science in Finance

- Work Experience

- Commodity Price Reporter
- Software Support Analyst at a maritime software company
- Data & Insights Specialist at a shipping company

Fake News

Problems &
Stakeholders



What is fake news

Misinformation

False and inaccurate information

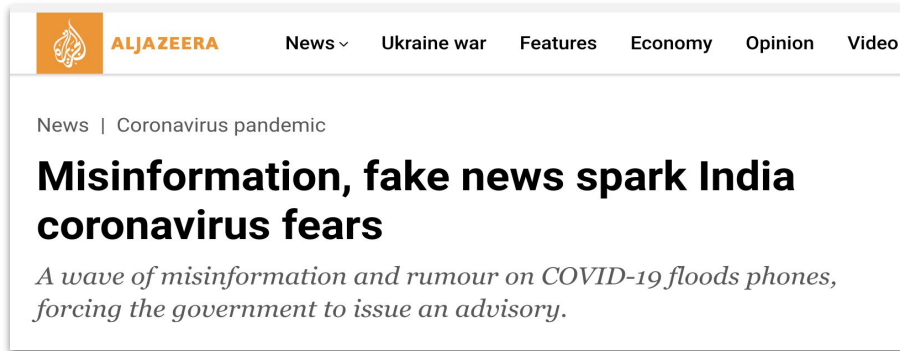
→ Misunderstanding of facts

Disinformation

Spread wrong information in bad faith

→ Propaganda, for political benefit

Global chaos from pandemic fake news



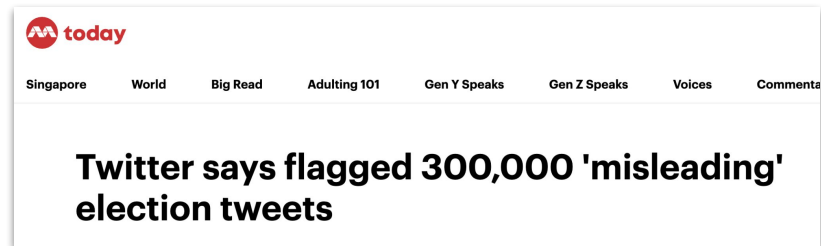
Unfounded fears of coronavirus



Geopolitical tensions

SOCIAL MEDIA COMPANIES

Taking steps to tackle fake news



Source: CNBC, TODAY

STAKEHOLDERS

**SOCIAL NETWORKING COMPANIES
THAT CIRCULATE NEWS ARTICLES**

Social networking companies benefit from using a robust fake news detection model

**DETECTS FAKE
NEWS EASILY**

✓ Save time verifying news

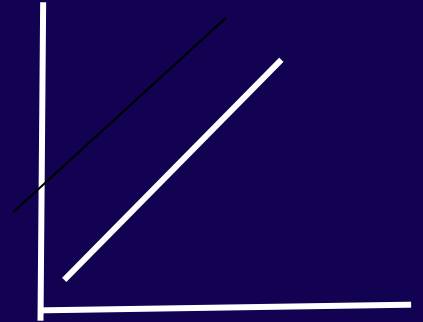
**REMOVES FAKE
NEWS FAST**

✓ Minimise spread of fake news

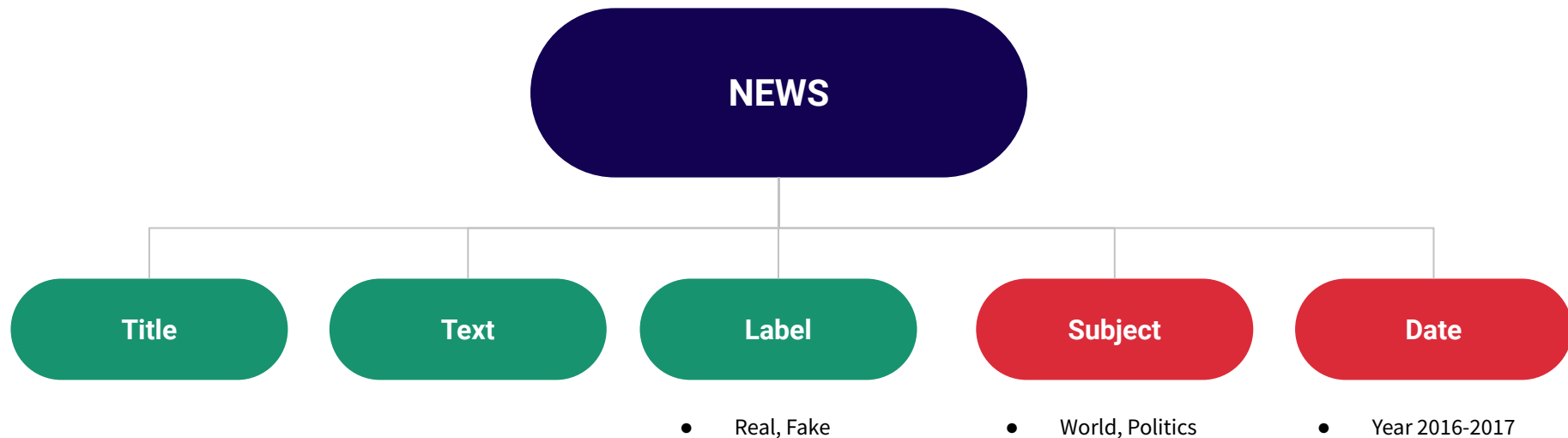
**IMPROVES
COMPANY
REPUTATION**

✓ Gain trust from the public
✓ Increase advertising business deals
✓ Reduce complications with regulators

Exploratory Data Analysis



Data Set



Data set: Kaggle

<https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset>

EDA

Number of real and fake news



About equal number of real (51%) and fake news (49%)

EDA

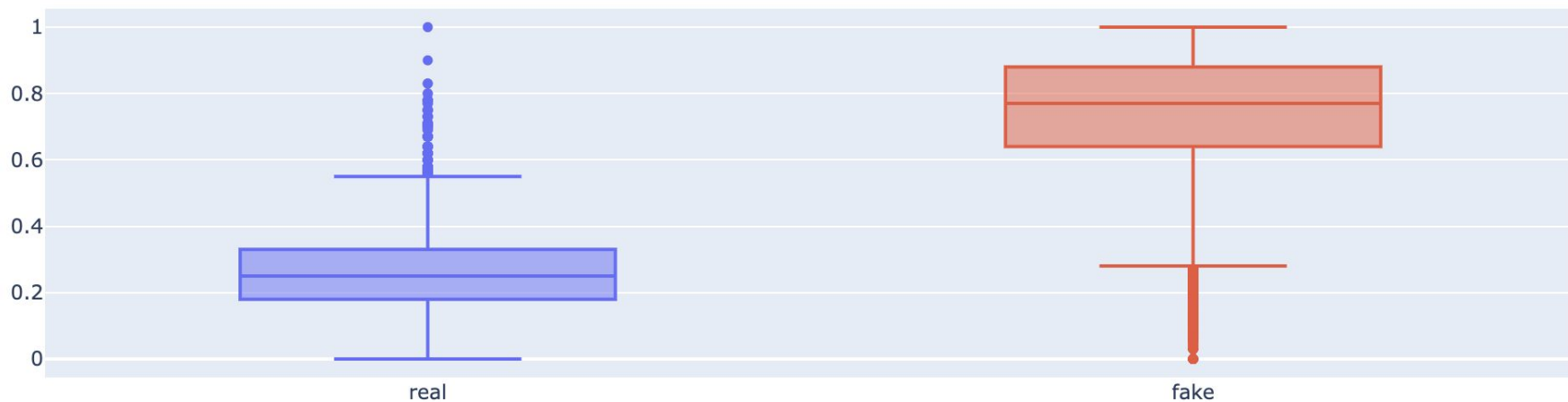
Title length in real and fake news



Real news has shorter title length than fake news

EDA

News title: Ratio of words in title case



Fake news titles tends to use title cases

EDA

News Title

Real News



Fake News



34% of fake news has the word 'video' in their titles

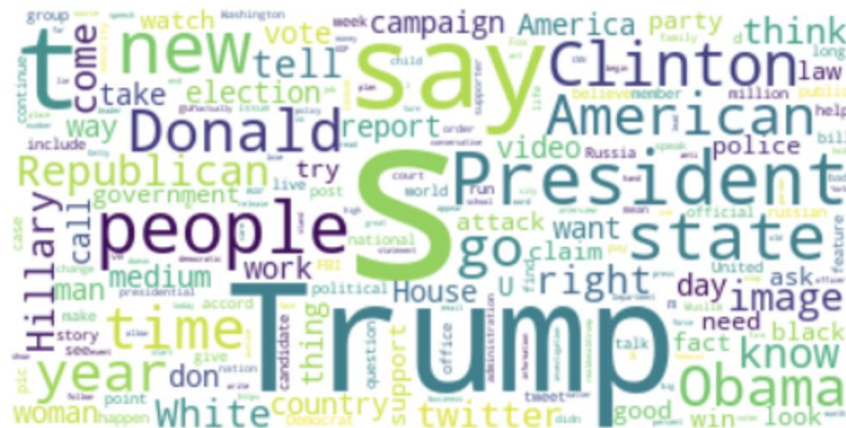
EDA

News Text

Real News



Fake News



'Reuters' shows up frequently in real news

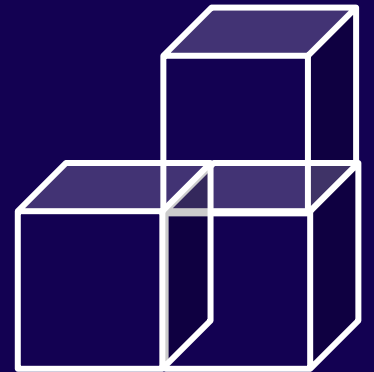
EDA

99%

of real news data are from Reuters

	title	text	real_or_fake
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	real
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	real
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	real
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	real
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	real

Modelling Results



Feature Selection and Models

Feature Selection

- CountVectorizer
- TF-IDF (word)
- TF-IDF (n-gram)

Supervised learning models

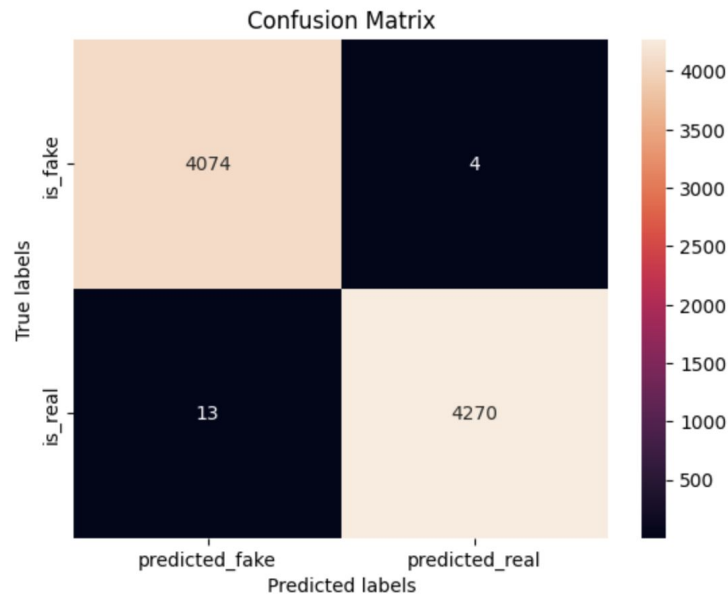
- Naive Bayes
- Support Vector Machine (SVM)
- Logistic Regression (LR)

Results

	Count Vectors	WordLevel TF-IDF	N-Gram Vectors
Naïve Bayes	0.962086	0.941514	0.974525
SVM	0.997847	0.997010	0.992106
LR	0.997967	0.991747	0.986963

Results show very high accuracy of 99.8%

Results



Only 0.1% of fake news incorrectly classified

Removing biases

Title

Remove word 'video'

Text

Remove beginning section of text with Reuters name

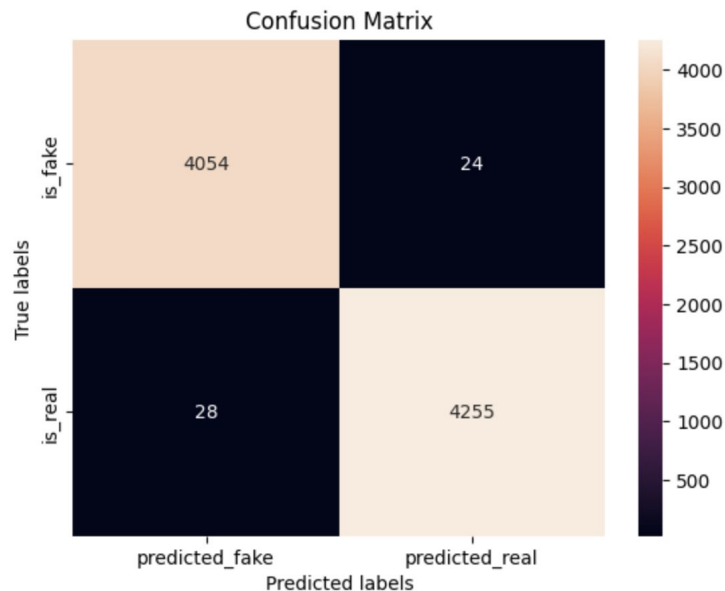
	title	text	real_or_fake
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	real

Results - without bias

	Count Vectors	WordLevel TF-IDF	N-Gram Vectors
Naïve Bayes	0.957302	0.937567	0.971654
SVM	0.989953	0.993781	0.988279
LR	0.991867	0.988757	0.984212

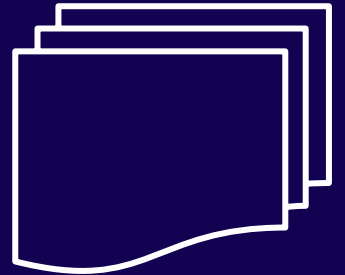
Results show very high accuracy of 99.4%

Results - without bias



Only 0.6% of fake news incorrectly classified

Conclusion



KEY TAKEAWAY

**Machine learning models
can be useful for social
media companies to
detect fake news**

FUTURE WORK

**Collect more data from
different news sources for
better representation of
available news**

End

Appendix

News links of screenshots used in slides:

<https://www.aljazeera.com/news/2020/3/10/misinformation-fake-news-spark-india-coronavirus-fears>

<https://www.bbc.com/news/world-us-canada-51599009>

<https://www.cnbc.com/2020/12/03/facebook-to-remove-misinformation-about-coronavirus-vaccines.html>

<https://www.todayonline.com/world/twitter-says-flagged-300000-misleading-election-tweets>