

# 'You're fake news'

A photograph of Donald Trump speaking at a podium. He is pointing his right index finger directly at the camera. He is wearing a dark suit jacket, a white shirt, and a dark tie. A black microphone is positioned in front of his mouth. The background consists of several American flags. The text 'You're fake news' is overlaid at the top of the image in a large, white, sans-serif font.

Classifying Fake  
News with Machine  
Learning

GA Data Science  
Part-Time

# Project Overview

Where did I get the data: <https://kaggle.com/fake-and-real-news-dataset>



## What is the problem?

- Fake news undermines legitimate and established systems, and encourages dissemination of false information.
- Fake news can be hard to identify
- **Is there a data-driven solution to this problem?**
- Can we build a machine-learning model that can successfully classify news articles as fake or real?

## What is the Modelling Strategy?

- I chose to deploy several classification models, which included Naive Bayes, Support Vector Machine and Logistic Regression, and all of which performed well.
- Ultimately, after processing the text with a tfidf vectorizer, the SVM model performed best with a test accuracy of 0.98%.

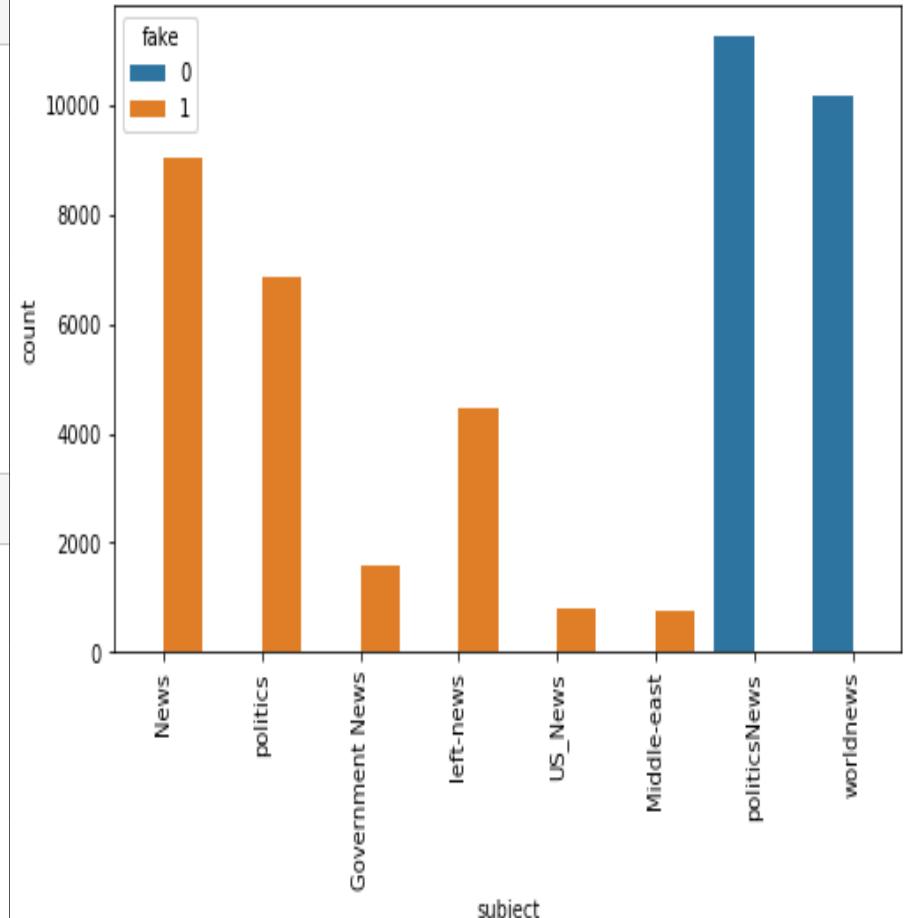
# What does the data look like?

news.head()

	title	text	subject	date	target
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn't wish all Americans ...	News	December 31, 2017	fake
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	fake
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	fake
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	fake
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	fake

news.tail()

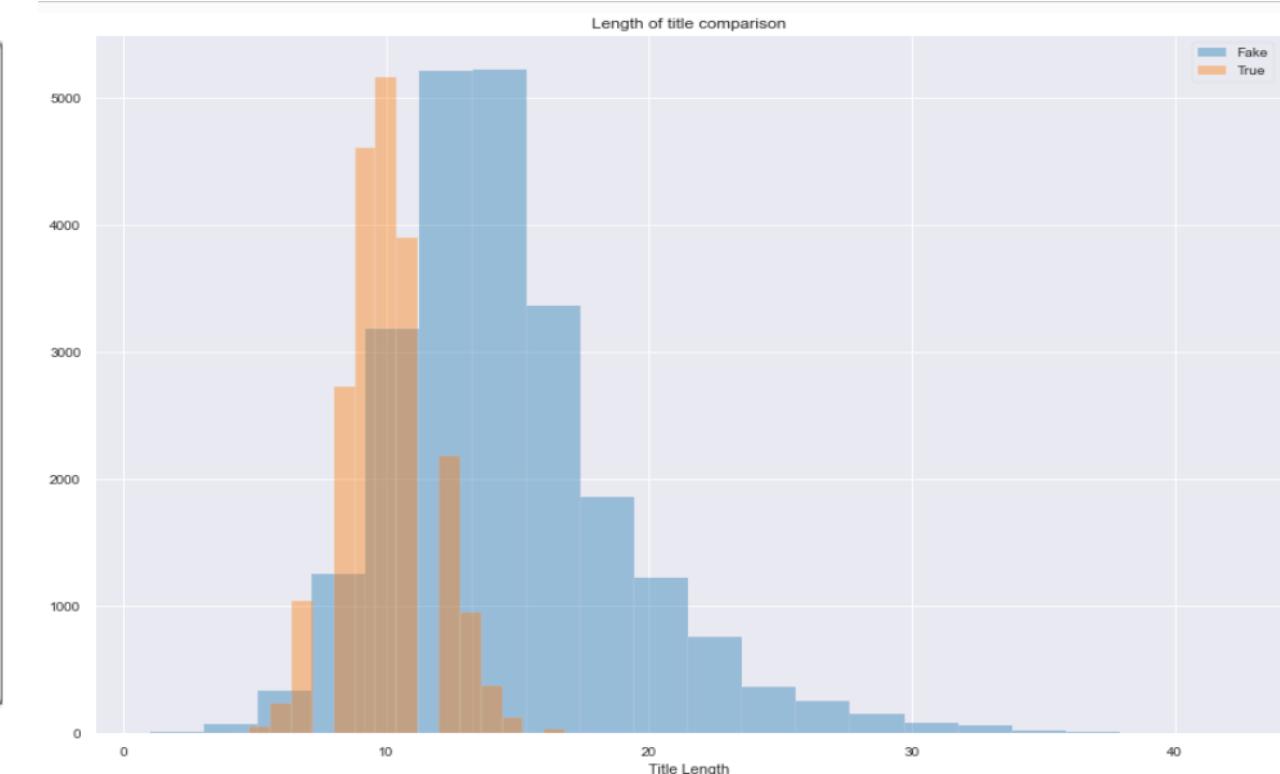
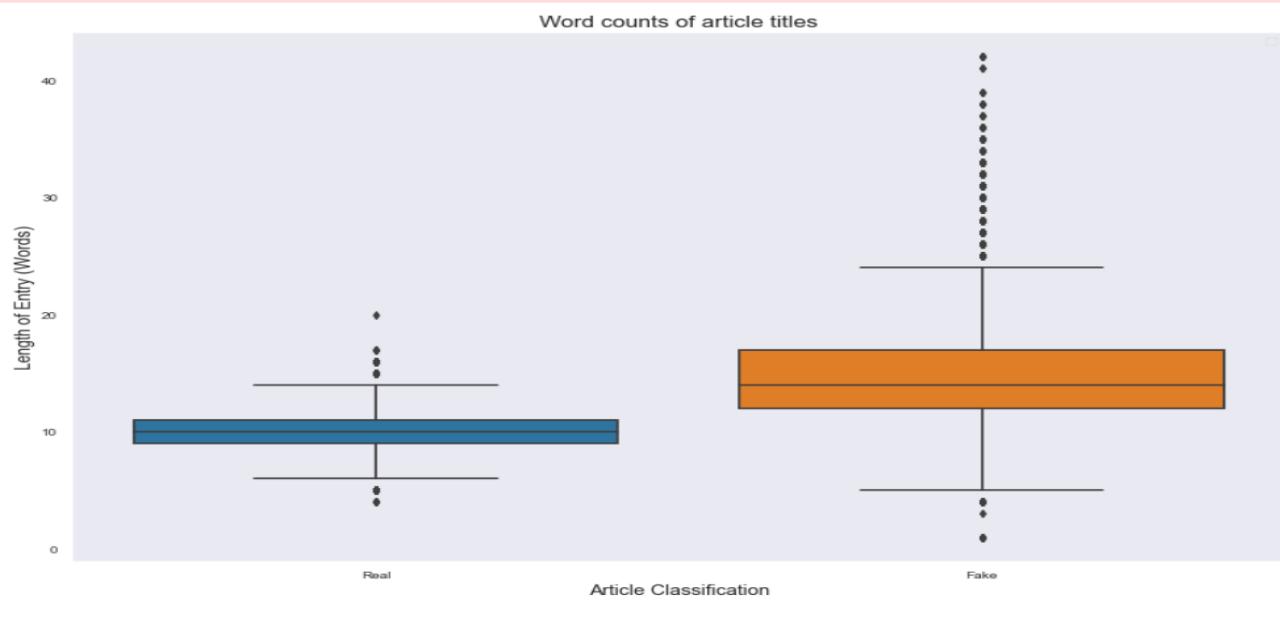
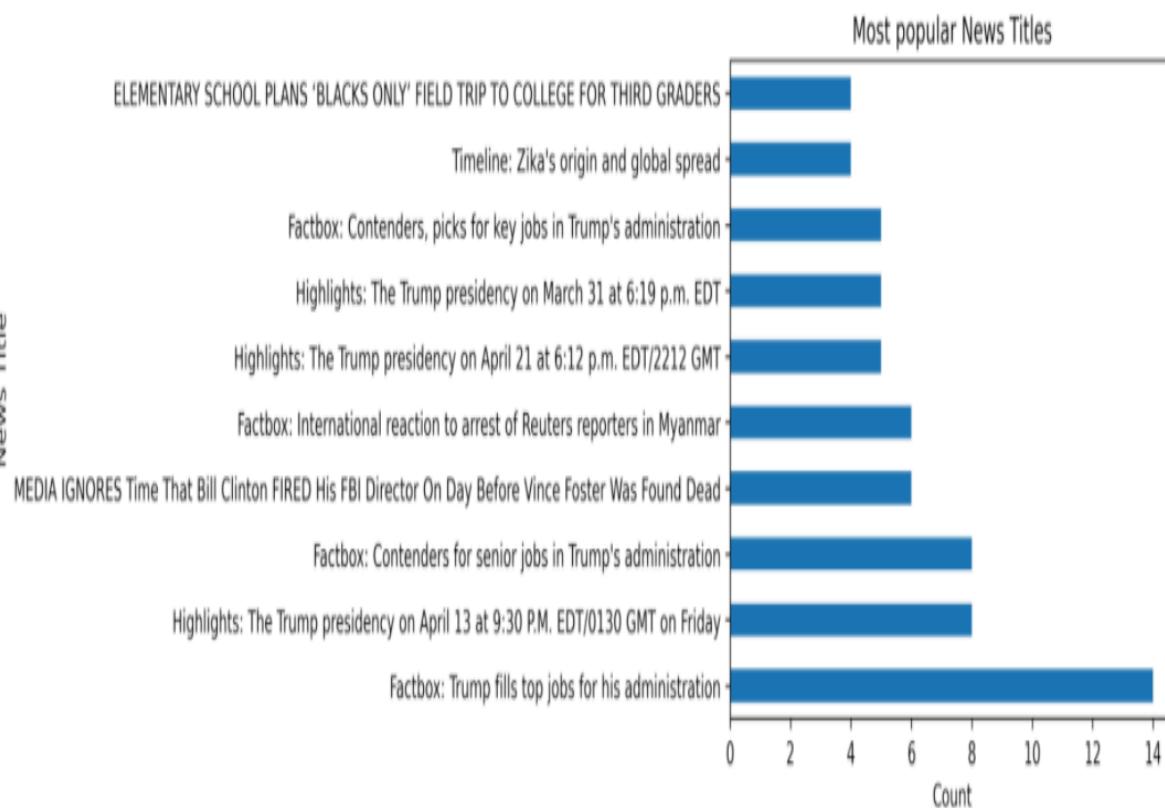
	title	text	subject	date	target
44893	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	true
44894	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of l...	worldnews	August 22, 2017	true
44895	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	true
44896	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	true
44897	Indonesia to buy \$1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	true



It is assumed that fake news articles will have one of the following characteristics:

- Article title will address the reader directly
- Sources will use very specific figures and numbers in the text body as well as the title
- Fake news will make alarmist claims using strong adjectives

E  
D  
A



# What are fake news stories saying?

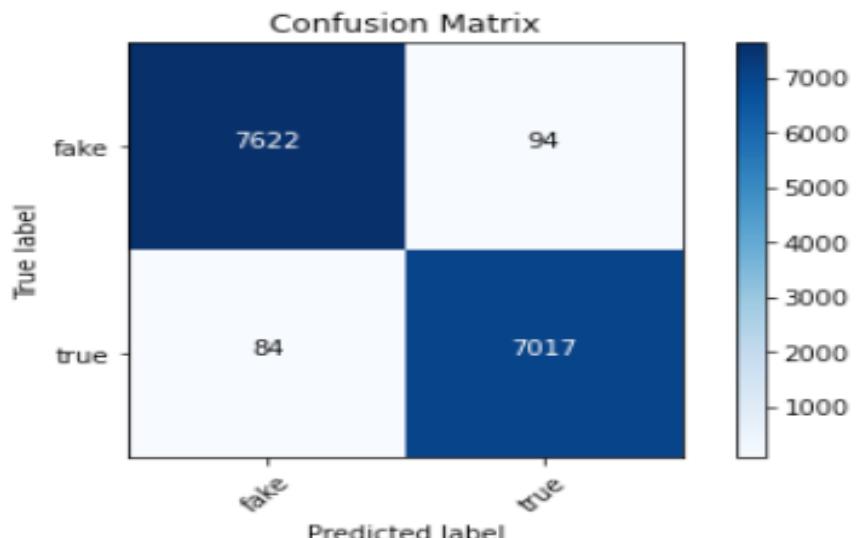
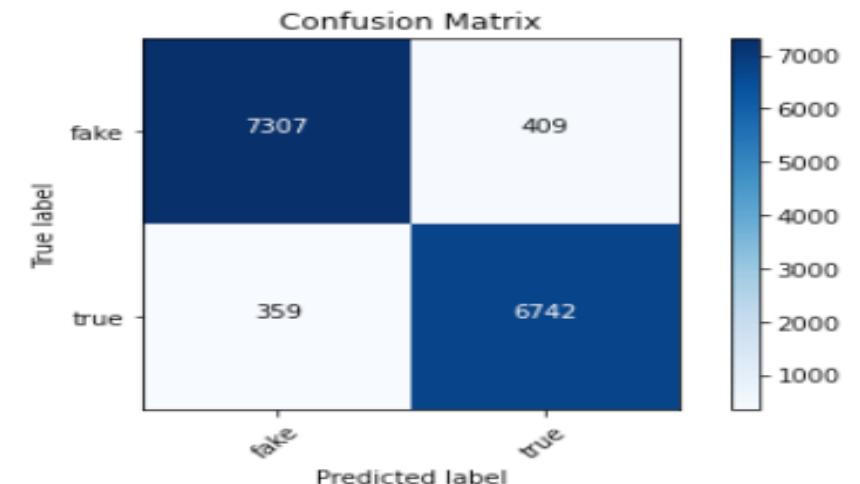
**FAKE**

# Modelling Insights

MNB

Model	Accuracy	Parameters
Multinomial Naïve Bayes Classifier (MNB) with Count Vectorizer	95.21%	Optimum score with 0.5 alphas
Multinomial Naïve Bayes Classifier (MNB) with TfIdf Vectorizer	94.82%	Optimum score with 0.2 alphas
Linear Support Vector Machine (SVM) with ML Pipeline	98.83%	Ran best with the TfIdf Vectorizer, optimum K = 1000, and GridSearch CV best score 99% accuracy
Logistic Regression with ML Pipeline	97.76%	CV = 5, C=10.0, Max iterations = 200

SVM



grid\_search.cv\_results\_

```
{'mean_fit_time': array([5.57449794, 5.25479801, 6.07305042]),
 'std_fit_time': array([0.48067196, 0.02016892, 0.80188844]),
 'mean_score_time': array([2.45098829, 2.41130702, 2.76148391]),
 'std_score_time': array([0.06553305, 0.02918204, 0.15165873]),
 'param_selector_k': masked_array(data=[50, 100, 1000],
 mask=[False, False, False]),
```