This jupyter notebook is prepared by Matthias Rathbun.

1. Load Data and perform basic EDA

1. import libraries necessary libraries and perform necessariy nltk download operations
2. As part of understanding how the columns are separated, read the file using the open function and create a list and show the first 10 items in the list
3. Based on your observation on how the data are separated, load the data set into pandas data frame and show the first 5 and last 5 rows
4. See whether there are any null values and remove all the rows with any null values, and then show again that there are no more null values
5. Generate a counterplot to show the number of news in each subject
6. Generate a counterplot to show the number of news in each category (fake/ True)
7. Generate two word clouds, one for fake news and one for true news, and observe the most frequent words in each category and just write your observation on them.
8. Create a column "AllText" that has the concatenated subject, title, and text [For example, for each news we have the subject, title, and text. We want a column that has all of this together as a large string] [See the example answer with the majority vote in this Link]
9. Using the dataframe's copy function, save the data frame into another dataframe so that you can use it later
10. Drop the title, text, subject, and date columns from the data frame as we will not use them separately. We have all the text in a single column that you have generated above
11. Calculate the length of each text (I mean AllText column) and put them in a length column
12. Plot two histograms to see the distribution of the lengths. One for fake news and one for true news. Write in words about the plots
13. Write in Words: What is TFIDF? How to create bag of words using sklearn? And how to generate TFIDF for the bag of words?

In [15]:
```
import pandas as pd
import nltk
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from nltk.stem import WordNetLemmatizer
from nltk.corpus import stopwords
from nltk.stem import PorterStemmer
from nltk.stem import LancasterStemmer
import seaborn as sns
from sklearn.model_selection import train_test_split
import string
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.naive_bayes import MultinomialNB
from sklearn.pipeline import Pipeline
from sklearn.metrics import classification_report, confusion_matrix
from newspaper import Article
from sklearn.neural_network import MLPClassifier
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
from tensorflow.keras.models import Sequential
from tensorflow.keras.optimizers import SGD
import numpy as np
from hyperas import optim
from hyperas.distributions import choice, uniform
from hyperopt import Trials, STATUS_OK, tpe
import _locale
_locale._getdefaultlocale = (lambda *args: ['en_US', 'utf8'])
%matplotlib inline
```

In [2]:
```
print("Num GPUs Available: ", len(tf.config.list_physical_devices('GPU')))
```
```
Num GPUs Available:  1
```

In [3]:
```
nltk.download('wordnet') # we will use wordnet lemmatizer
nltk.download('averaged_perceptron_tagger') #to be used for tagger
nltk.download('punkt')
nltk.download('stopwords')
```
```
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\matth\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     C:\Users\matth\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\matth\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\matth\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```
Out[3]: True

In [4]:
```
f = open("news.csv", "r", encoding="utf8")
i = 0
for x in f:
    print(x)
    i+=1
    if i > 9:
        break
```

```
title	text	subject	date	target

As U.S. budget fight looms, Republicans flip their fiscal script	WASHINGTON (Reuters) - The head of a conservative Republican faction in the U.S. Congress, who voted this month for a huge expansion of the national debt to pay for tax cuts, called himself a “fiscal conservative” on Sunday and urged budget restraint in 2018. In keeping with a sharp pivot under way among Republicans, U.S. Representative Mark Meadows, speaking on CBS’ “Face the Nation,” drew a hard line on federal spending, which lawmakers are bracing to do battle over in January. When they return from the holidays on Wednesday, lawmakers will begin trying to pass a federal budget in a fight likely to be linked to other issues, such as immigration policy, even as the November congressional election campaigns approach in which Republicans will seek to keep control of Congress. President Donald Trump and his Republicans want a big budget increase in military spending, while Democrats also want proportional increases for non-defense “discretionary” spending on programs that support education, scientific research, infrastructure, public health and environmental protection. “The (Trump) administration has already been willing to say: ‘We’re going to increase non-defense discretionary spending ... by about 7 percent,’” Meadows, chairman of the small but influential House Freedom Caucus, said on the program. “Now, Democrats are saying that’s not enough, we need to give the government a pay raise of 10 to 11 percent. For a fiscal conservative, I don’t see where the rationale is. ... Eventually you run out of other people’s money,” he said. Meadows was among Republicans who voted in late December for their party’s debt-financed tax overhaul, which is expected to balloon the federal budget deficit and add about $1.5 trillion over 10 years to the $20 trillion national debt. “It’s interesting to hear Mark talk about fiscal responsibility,” Democratic U.S. Representative Joseph Crowley said on CBS. Crowley said the Republican tax bill would require the  United States to borrow $1.5 trillion, to be paid off by future generations, to finance tax cuts for corporations and the rich. “This is one of the least ... fiscally responsible bills we’ve ever seen passed in the history of the House of Representatives. I think we’re going to be paying for this for many, many years to come,” Crowley said. Republicans insist the tax package, the biggest U.S. tax overhaul in more than 30 years,  will boost the economy and job growth. House Speaker Paul Ryan, who also supported the tax bill, recently went further than Meadows, making clear in a radio interview that welfare or “entitlement reform,” as the party often calls it, would be a top Republican priority in 2018. In Republican parlance, “entitlement” programs mean food stamps, housing assistance, Medicare and Medicaid health insurance for the elderly, poor and disabled, as well as other programs created by Washington to assist the needy. Democrats seized on Ryan’s early December remarks, saying they showed Republicans would try to pay for their tax overhaul by seeking spending cuts for social programs. But the goals of House Republicans may have to take a back seat to the Senate, where the votes of some Democrats will be needed to approve a budget and prevent a government shutdown. Democrats will use their leverage in the Senate, which Republicans narrowly control, to defend both discretionary non-defense programs and social spending, while tackling the issue of the “Dreamers,” people brought illegally to the country as children. Trump in September put a March 2018 expiration date on the Deferred Action for Childhood Arrivals, or DACA, program, which protects the young immigrants from deportation and provides them with work permits. The president has said in recent Twitter messages he wants funding for his proposed Mexican border wall and other immigration law changes in exchange for agreeing to help the Dreamers. Representative Debbie Dingell told CBS she did not favor linking that issue to other policy objectives, such as wall funding. “We need to do DACA clean,” she said.  On Wednesday, Trump aides will meet with congressional leaders to discuss those issues. That will be followed by a weekend of strategy sessions for Trump and Republican leaders on Jan. 6 and 7, the White House said. Trump was also scheduled to meet on Sunday with Florida Republican Governor Rick Scott, who wants more emergency aid. The House has passed an $81 billion aid package after hurricanes in Florida, Texas and Puerto Rico, and wildfires in California. The package far exceeded the $44 billion requested by the Trump administration. The Senate has not yet voted on the aid.	politicsNews	December 31, 2017	1

U.S. military to accept transgender recruits on Monday: Pentagon	WASHINGTON (Reuters) - Transgender people will be allowed for the first time to enlist in the U.S. military starting on Monday as ordered by federal courts, the Pentagon said on Friday, after President Donald Trump’s administration decided not to appeal rulings that blocked his transgender ban. Two federal appeals courts, one in Washington and one in Virginia, last week rejected the administration’s request to put on hold orders by lower court judges requiring the military to begin accepting transgender recruits on Jan. 1. A Justice Department official said the administration will not challenge those rulings. “The Department of Defense has announced that it will be releasing an independent study of these issues in the coming weeks. So rather than litigate this interim appeal before that occurs, the administration has decided to wait for DOD’s study and will continue to defend the president’s lawful authority in District Court in the meantime,” the official said, speaking on condition of anonymity. In September, the Pentagon said it had created a panel of senior officials to study how to implement a directive by Trump to prohibit transgender individuals from serving. The Defense Department has until Feb. 21 to submit a plan to Trump. Lawyers representing currently-serving transgender service members and aspiring recruits said they had expected the administration to appeal the rulings to the conservative-majority Supreme Court, but were hoping that would not happen. Pentagon spokeswoman Heather Babb said in a statement: “As mandated by court order, the Department of Defense is prepared to begin accessing transgender applicants for military service Jan. 1. All applicants must meet all accession standards.” Jennifer Levi, a lawyer with gay, lesbian and transgender advocacy group GLAD, called the decision not to appeal “great news.” “I’m hoping it means the government has come to see that there is no way to justify a ban and that it’s not good for the military or our country,” Levi said. Both GLAD and the American Civil Liberties Union represent plaintiffs in the lawsuits filed against the administration. In a move that appealed to his hard-line conservative supporters, Trump announced in July that he would prohibit transgender people from serving in the military, reversing Democratic President Barack Obama’s policy of accepting them. Trump said on Twitter at the time that the military “cannot be burdened with the tremendous medical costs and disruption that transgender in the military would entail.” Four federal judges - in Baltimore, Washington, D.C., Seattle and Riverside, California - have issued rulings blocking Trump’s ban while legal challenges to the Republican president’s policy proceed. The judges said the ban would likely violate the right under the U.S. Constitution to equal protection under the law. The Pentagon on Dec. 8 issued guidelines to recruitment personnel in order to enlist transgender applicants by Jan. 1. The memo outlined medical requirements and specified how the applicants’ sex would be identified and even which undergarments they would wear. The Trump administration previously said in legal papers that the armed forces were not prepared to train thousands of personnel on the medical standards needed to process transgender applicants and might have to accept “some individuals who are not medically fit for service.” The Obama administration had set a deadline of July 1, 2017, to begin accepting transgender recruits. But Trump’s defense secretary, James Mattis, postponed that date to Jan. 1, 2018, which the president’s ban then put off indefinitely. Trump has taken other steps aimed at rolling back transgender rights. In October, his administration said a federal law banning gender-based workplace discrimination does not protect transgender employees, reversing another Obama-era position. In February, Trump rescinded guidance issued by the Obama administration saying that public schools should allow transgender students to use the restroom that corresponds to their gender identity.	politicsNews	December 29, 2017	1

Senior U.S. Republican senator: 'Let Mr. Mueller do his job'	WASHINGTON (Reuters) - The special counsel investigation of links between Russia and President Trump’s 2016 election campaign should continue without interference in 2018, despite calls from some Trump administration allies and Republican lawmakers to shut it down, a prominent Republican senator said on Sunday. Lindsey Graham, who serves on the Senate armed forces and judiciary committees, said Department of Justice Special Counsel Robert Mueller needs to carry on with his Russia investigation without political interference. “This investigation will go forward. It will be an investigation conducted without political influence,” Graham said on CBS’s Face the Nation news program. “And we all need to let Mr. Mueller do his job. I think he’s the right guy at the right time.”  The question of how Russia may have interfered in the election, and how Trump’s campaign may have had links with or co-ordinated any such effort, has loomed over the White House since Trump took office in January. It shows no sign of receding as Trump prepares for his second year in power, despite intensified rhetoric from some Trump allies in recent weeks accusing Mueller’s team of bias against the Republican president. Trump himself seemed to undercut his supporters in an interview last week with the New York Times in which he said he expected Mueller was “going to be fair.”    Russia’s role in the election and the question of possible links to the Trump campaign are the focus of multiple inquiries in Washington. Three committees of the Senate and the House of Representatives are investigating, as well as Mueller, whose team in May took over an earlier probe launched by the U.S. Federal Bureau of Investigation (FBI). Several members of the Trump campaign and administration have been convicted or indicted in the investigation.  Trump and his allies deny any collusion with Russia during the campaign, and the Kremlin has denied meddling in the election. Graham said he still wants an examination of the FBI’s use of a dossier on links between Trump and Russia that was compiled by a former British spy, Christopher Steele, which prompted Trump allies and some Republicans to question Mueller’s inquiry.   On Saturday, the New York Times reported that it was not that dossier that triggered an early FBI probe, but a tip from former Trump campaign foreign policy adviser George Papadopoulos to an Australian diplomat that Russia had damaging information about former Trump rival Hillary Clinton.  “I want somebody to look at the way the Department of Justice used this dossier. It bothers me greatly the way they used it, and I want somebody to look at it,” Graham said. But he said the Russia investigation must continue. “As a matter of fact, it would hurt us if we ignored it,” he said.	politicsNews	December 31, 2017	1

FBI Russia probe helped by Australian diplomat tip-off: NYT	WASHINGTON (Reuters) - Trump campaign adviser George Papadopoulos told an Australian diplomat in May 2016 that Russia had political dirt on Democratic presidential candidate Hillary Clinton, the New York Times reported on Saturday. The conversation between Papadopoulos and the diplomat, Alexander Downer, in London was a driving factor behind the FBI’s decision to open a counter-intelligence investigation of Moscow’s contacts with the Trump campaign, the Times reported. Two months after the meeting, Australian officials passed the information that came from Papadopoulos to their American counterparts when leaked Democratic emails began appearing online, according to the newspaper, which cited four current and former U.S. and foreign officials. Besides the information from the Australians, the probe by the Federal Bureau of Investigation was also propelled by intelligence from other friendly governments, including the British and Dutch, the Times said. Papadopoulos, a Chicago-based international energy lawyer, pleaded guilty on Oct. 30 to lying to FBI agents about contacts with people who claimed to have ties to top Russian officials. It was the first criminal charge alleging links between the Trump campaign and Russia. The White House has played down the former aide’s campaign role, saying it was “extremely limited” and that any actions he took would have been on his own. The New York Times, however, reported that Papadopoulos helped set up a meeting between then-candidate Donald Trump and Egyptian President Abdel Fattah al-Sisi and edited the outline of Trump’s first major foreign policy speech in April 2016. The federal investigation, which is now being led by Special Counsel Robert Mueller, has hung over Trump’s White House since he took office almost a year ago. Some Trump allies have recently accused Mueller’s team of being biased against the Republican president. Lawyers for Papadopoulos did not immediately respond to requests by Reuters for comment. Mueller’s office declined to comment. Trump’s White House attorney, Ty Cobb, declined to comment on the New York Times report. “Out of respect for the special counsel and his process, we are not commenting on matters such as this,” he said in a statement. Mueller has charged four Trump associates, including Papadopoulos, in his investigation. Russia has denied interfering in the U.S. election and Trump has said there was no collusion between his campaign and Moscow.	politicsNews	December 30, 2017	1

Trump wants Postal Service to charge 'much more' for Amazon shipments	SEATTLE/WASHINGTON (Reuters) - President Donald Trump called on the U.S. Postal Service on Friday to charge “much more” to ship packages for Amazon (AMZN.O), picking another fight with an online retail giant he has criticized in the past.     “Why is the United States Post Office, which is losing many billions of dollars a year, while charging Amazon and others so little to deliver their packages, making Amazon richer and the Post Office dumber and poorer? Should be charging MUCH MORE!” Trump wrote on Twitter.  The president’s tweet drew fresh attention to the fragile finances of the Postal Service at a time when tens of millions of parcels have just been shipped all over the country for the holiday season.  The U.S. Postal Service, which runs at a big loss, is an independent agency within the federal government and does not receive tax dollars for operating expenses, according to its website.  Package delivery has become an increasingly important part of its business as the Internet has led to a sharp decline in the amount of first-class letters. The president does not determine postal rates. They are set by the Postal Regulatory Commission, an independent government agency with commissioners selected by the president from both political parties. That panel raised prices on packages by almost 2 percent in November.  Amazon was founded by Jeff Bezos, who remains the chief executive officer of the retail company and is the richest person in the world, according to Bloomberg News. Bezos also owns The Washington Post, a newspaper Trump has repeatedly railed against in his criticisms of the news media. In tweets over the past year, Trump has said the “Amazon Washington Post” fabricated stories. He has said Amazon does not pay sales tax, which is not true, and so hurts other retailers, part of a pattern by the former businessman and reality television host of periodically turning his ire on big American companies since he took office in January. Daniel Ives, a research analyst at GBH Insights, said Trump’s comment could be taken as a warning to the retail giant. However, he said he was not concerned for Amazon. “We do not see any price hikes in the future. However, that is a risk that Amazon is clearly aware of and (it) is building out its distribution (system) aggressively,” he said. Amazon has shown interest in the past in shifting into its own delivery service, including testing drones for deliveries. In 2015, the company spent $11.5 billion on shipping, 46 percent of its total operating expenses that year.  Amazon shares were down 0.86 percent to $1,175.90 by early afternoon. Overall, U.S. stock prices were down slightly on Friday.  Satish Jindel, president of ShipMatrix Inc, which analyzes shipping data, disputed the idea that the Postal Service charges less than United Parcel Service Inc (UPS.N) and FedEx Corp (FDX.N), the other biggest players in the parcel delivery business in the United States. Many customers get lower rates from UPS and FedEx than they would get from the post office for comparable services, he said. The Postal Service delivers about 62 percent of Amazon packages, for about 3.5 to 4 million a day during the current peak year-end holiday shipping season, Jindel said. The Seattle-based company and the post office have an agreement in which mail carriers take Amazon packages on the last leg of their journeys, from post offices to customers’ doorsteps. Amazon’s No. 2 carrier is UPS, at 21 percent, and FedEx is third, with 8 percent or so, according to Jindel. Trump’s comment tapped into a debate over whether Postal Service pricing has kept pace with the rise of e-commerce, which has flooded the mail with small packages.Private companies like UPS have long claimed the current system unfairly undercuts their business. Steve Gaut, a spokesman for UPS, noted that the company values its “productive relationship” with the postal service, but that it has filed with the Postal Regulatory Commission its concerns about the postal service’s methods for covering costs. Representatives for Amazon, the White House, the U.S. Postal Service and FedEx declined comment or were not immediately available for comment on Trump’s tweet. According to its annual report, the Postal Service lost $2.74 billion this year, and its deficit has ballooned to $61.86 billion.  While the Postal Service’s revenue for first class mail, marketing mail and periodicals is flat or declining, revenue from package delivery is up 44 percent since 2014 to $19.5 billion in the fiscal year ended Sept. 30, 2017. But it also lost about $2 billion in revenue when a temporary surcharge expired in April 2016. According to a Government Accountability Office report in February, the service is facing growing personnel expenses, particularly $73.4 billion in unfunded pension and benefits liabilities. The Postal Service has not announced any plans to cut costs. By law, the Postal Service has to set prices for package delivery to cover the costs attributable to that service. But the postal service allocates only 5.5 percent of its total costs to its business of shipping packages even though that line of business is 28 percent of its total revenue.	politicsNews	December 29, 2017	1

White House, Congress prepare for talks on spending, immigration	WEST PALM BEACH, Fla./WASHINGTON (Reuters) - The White House said on Friday it was set to kick off talks next week with Republican and Democratic congressional leaders on immigration policy, government spending and other issues that need to be wrapped up early in the new year. The expected flurry of legislative activity comes as Republicans and Democrats begin to set the stage for midterm congressional elections in November. President Donald Trump’s Republican Party is eager to maintain control of Congress while Democrats look for openings to wrest seats away in the Senate and the House of Representatives. On Wednesday, Trump’s budget chief Mick Mulvaney and legislative affairs director Marc Short will meet with Senate Majority Leader Mitch McConnell and House Speaker Paul Ryan - both Republicans - and their Democratic counterparts, Senator Chuck Schumer and Representative Nancy Pelosi, the White House said. That will be followed up with a weekend of strategy sessions for Trump, McConnell and Ryan on Jan. 6 and 7 at the Camp David presidential retreat in Maryland, according to the White House. The Senate returns to work on Jan. 3 and the House on Jan. 8. Congress passed a short-term government funding bill last week before taking its Christmas break, but needs to come to an agreement on defense spending and various domestic programs by Jan. 19, or the government will shut down. Also on the agenda for lawmakers is disaster aid for people hit by hurricanes in Puerto Rico, Texas and Florida, and by wildfires in California. The House passed an $81 billion package in December, which the Senate did not take up. The White House has asked for a smaller figure, $44 billion. Deadlines also loom for soon-to-expire protections for young adult immigrants who entered the country illegally as children, known as “Dreamers.” In September, Trump ended Democratic former President Barack Obama’s Deferred Action for Childhood Arrivals (DACA) program, which protected Dreamers from deportation and provided work permits, effective in March, giving Congress until then to devise a long-term solution. Democrats, some Republicans and a number of large companies have pushed for DACA protections to continue. Trump and other Republicans have said that it will not happen without Congress approving broader immigration policy changes and tougher border security. Democrats oppose funding for a wall promised by Trump along the U.S.-Mexican border.  “The Democrats have been told, and fully understand, that there can be no DACA without the desperately needed WALL at the Southern Border and an END to the horrible Chain Migration & ridiculous Lottery System of Immigration etc,” Trump said in a Twitter post on Friday. Trump wants to overhaul immigration rules for extended families and others seeking to live in the United States. Republican U.S. Senator Jeff Flake, a frequent critic of the president, said he would work with Trump to protect Dreamers. “We can fix DACA in a way that beefs up border security, stops chain migration for the DREAMers, and addresses the unfairness of the diversity lottery. If POTUS (Trump) wants to protect these kids, we want to help him keep that promise,” Flake wrote on Twitter. Congress in early 2018 also must raise the U.S. debt ceiling to avoid a government default. The U.S. Treasury would exhaust all of its borrowing options and run dry of cash to pay its bills by late March or early April if Congress does not raise the debt ceiling before then, according to the nonpartisan Congressional Budget Office. Trump, who won his first major legislative victory with the passage of a major tax overhaul this month, has also promised a major infrastructure plan.	politicsNews	December 29, 2017	1

Trump says Russia probe will be fair, but timeline unclear: NYT	WEST PALM BEACH, Fla (Reuters) - President Donald Trump said on Thursday he believes he will be fairly treated in a special counsel investigation into Russian meddling in the U.S. presidential election, but said he did not know how long the probe would last. The federal investigation has hung over Trump’s White House since he took office almost a year ago, and some Trump allies have in recent weeks accused the team of Justice Department Special Counsel Robert Mueller of being biased against the Republican president. But in an interview with the New York Times, Trump appeared to shrug off concerns about the investigation, which was prompted by U.S. intelligence agencies’ conclusion that Russia tried to help Trump defeat Democrat Hillary Clinton by hacking and releasing embarrassing emails and disseminating propaganda. “There’s been no collusion. But I think he’s going to be fair,” Trump said in what the Times described as a 30-minute impromptu interview at his golf club in West Palm Beach, Florida. Mueller has charged four Trump associates in his investigation. Russia has denied interfering in the U.S. election. U.S. Deputy Attorney General Rod Rosenstein said this month that he was not aware of any impropriety by Mueller’s team. Trump’s lawyers have been saying for weeks that they had expected the Mueller investigation to wrap up quickly, possibly by the end of 2017. Mueller has not commented on how long it will last. Trump told the Times that he did not know how long the investigation would take. “Timing-wise, I can’t tell you. I just don’t know,” he said. Trump said he thought a prolonged probe “makes the country look bad” but said it has energized his core supporters. “What it’s done is, it’s really angered the base and made the base stronger. My base is strong than it’s ever been,” he said. The interview was a rare break in Trump’s Christmas vacation in Florida. He has golfed each day aside from Christmas Day, and mainly kept a low profile, apart from the occasional flurry of tweets. He spent one day golfing with Republican Senator David Perdue from Georgia, who has pushed legislation to cap immigration numbers, and had dinner on Thursday with Commerce Secretary Wilbur Ross, an international trade hawk. Trump told the Times he hoped to work with Democrats in the U.S. Congress on a spending plan to fix roads and other infrastructure, and on protections for a group of undocumented immigrants who were brought to the United States as children. Trump spoke about trade issues, saying he had backed off his hard line on Chinese trade practices in the hope that Beijing would do more to pressure North Korea to end its nuclear and missile testing program. He said he had been disappointed in the results. He also complained about the North American Free Trade Agreement (NAFTA), which his administration is attempting to renegotiate in talks with Mexico and Canada. Trump said Canadian Prime Minister Justin Trudeau had played down the importance of Canadian oil and lumber exports to the United States when looking at the balance of trade between the two countries. “If I don’t make the right deal, I’ll terminate NAFTA in two seconds. But we’re doing pretty good,” Trump said.	politicsNews	December 29, 2017	1

Factbox: Trump on Twitter (Dec 29) - Approval rating, Amazon	The following statements were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @POTUS.  The opinions expressed are his own. Reuters has not edited the statements or confirmed their accuracy.  @realDonaldTrump : - While the Fake News loves to talk about my so-called low approval rating, @foxandfriends just showed that my rating on Dec. 28, 2017, was approximately the same as President Obama on Dec. 28, 2009, which was 47%...and this despite massive negative Trump coverage & Russia hoax! [0746 EST] - Why is the United States Post Office, which is losing many billions of dollars a year, while charging Amazon and others so little to deliver their packages, making Amazon richer and the Post Office dumber and poorer? Should be charging MUCH MORE! [0804 EST] -- Source link: (bit.ly/2jBh4LU) (bit.ly/2jpEXYR)	politicsNews	December 29, 2017	1

Trump on Twitter (Dec 28) - Global Warming	The following statements were posted to the verified Twitter accounts of U.S. President Donald Trump, @realDonaldTrump and @POTUS.  The opinions expressed are his own. Reuters has not edited the statements or confirmed their accuracy.  @realDonaldTrump : - Together, we are MAKING AMERICA GREAT AGAIN! bit.ly/2lnpKaq [1814 EST] - In the East, it could be the COLDEST New Year’s Eve on record. Perhaps we could use a little bit of that good old Global Warming that our Country, but not other countries, was going to pay TRILLIONS OF DOLLARS to protect against. Bundle up! [1901 EST] -- Source link: (bit.ly/2jBh4LU) (bit.ly/2jpEXYR)	politicsNews	December 29, 2017	1
```

In [5]:
```
df = pd.read_csv("news.csv", sep = "\t")
```

In [6]:
```
df
```

Out[6]:

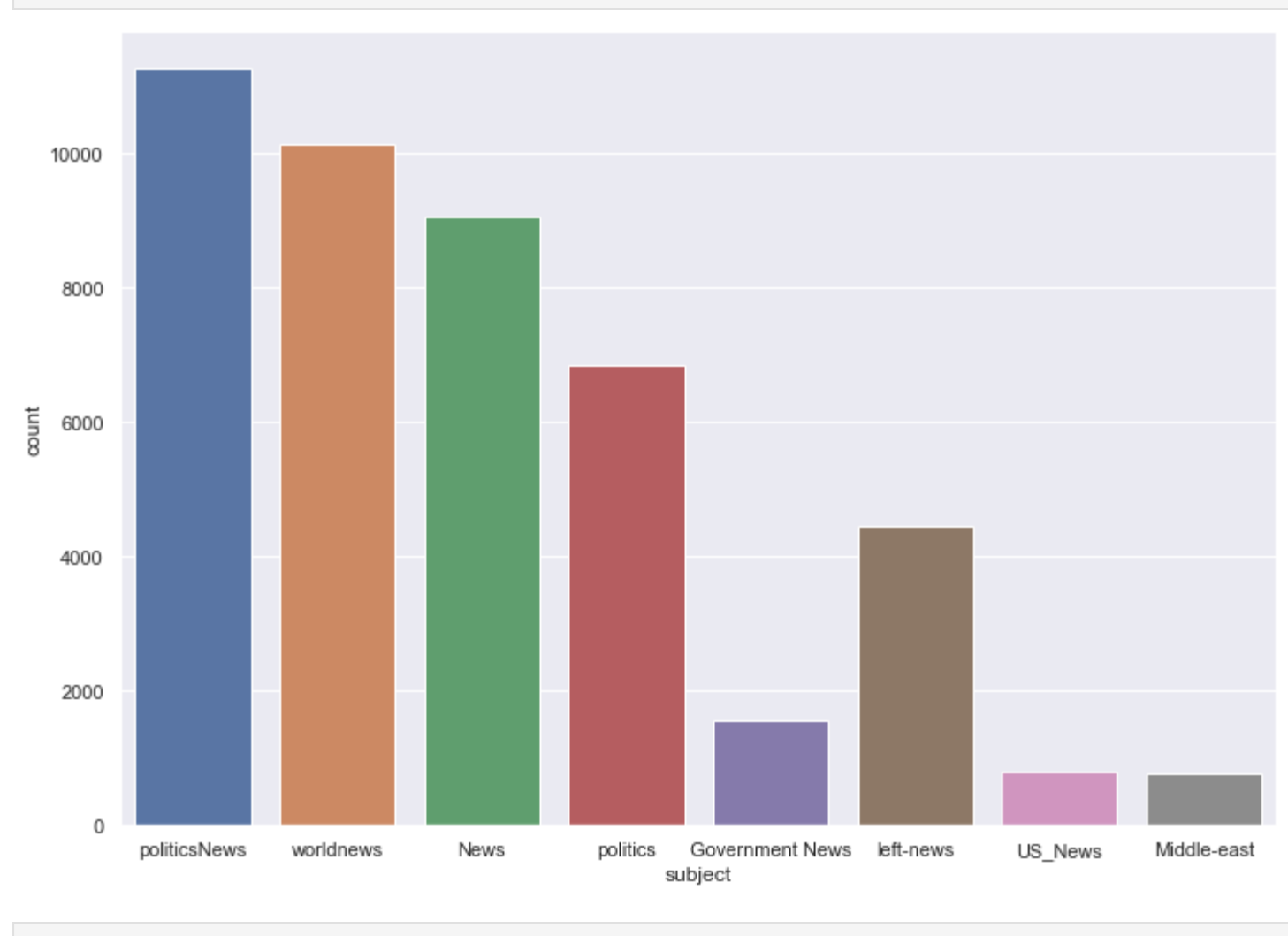
	title	text	subject	date	target
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	1
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	1
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	1
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	1
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	1
...	...	...	...	...	...
44893	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0
44894	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0
44895	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0
44896	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0
44897	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0

44898 rows × 5 columns

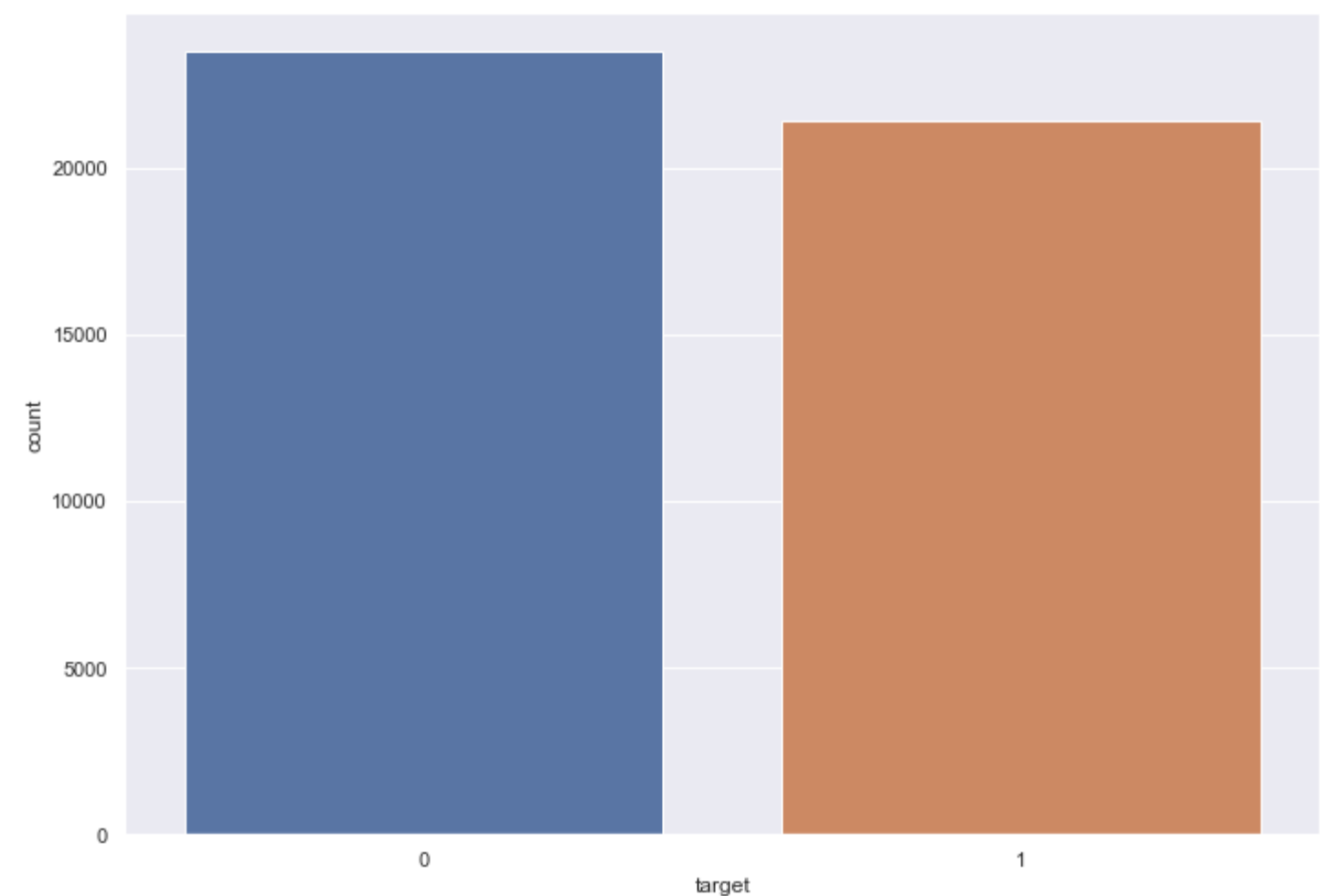
In [7]:
```
df.isna().sum().sort_values(ascending = False)
```

Out[7]:
```
title      0
text       0
subject    0
date       0
target     0
dtype: int64
```

In [8]:
```
sns.set(rc={'figure.figsize':(11.7,8.27)})
ax = sns.countplot(x="subject", data=df)
```



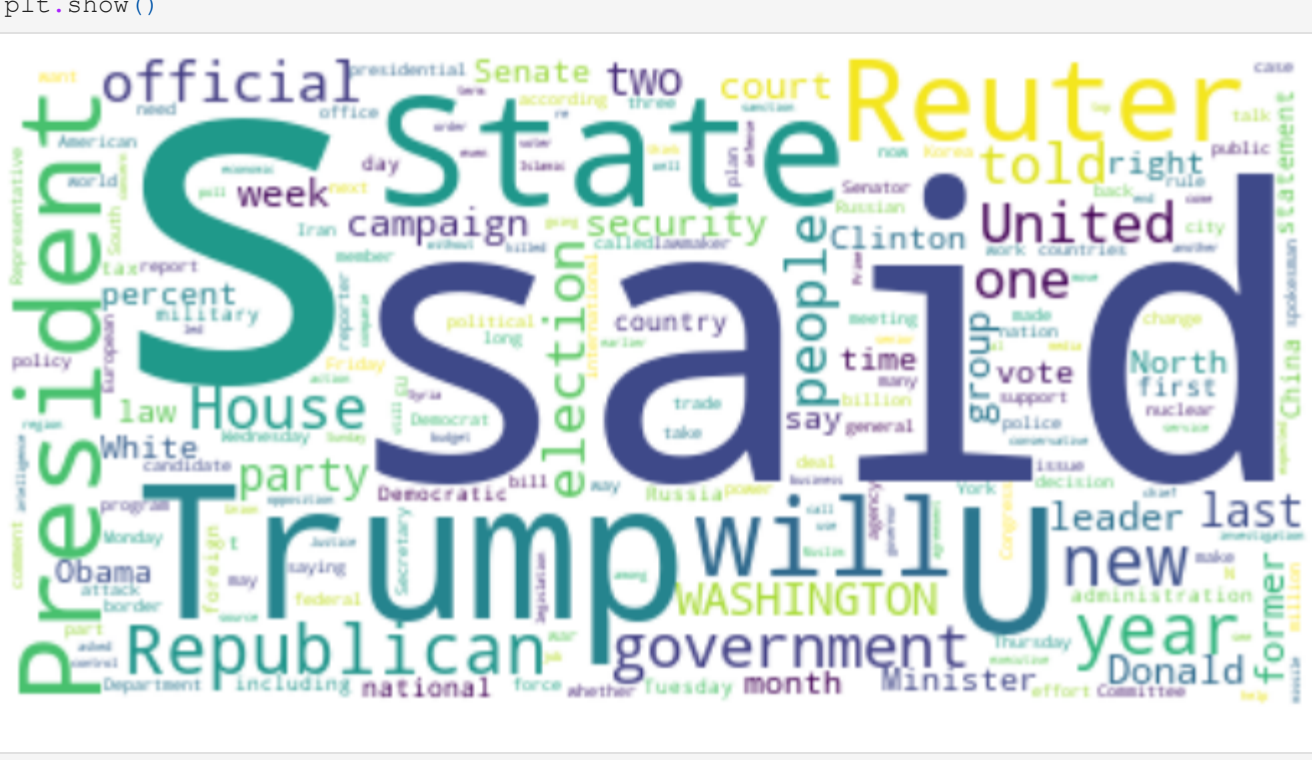
In [9]:
```
ax = sns.countplot(x="target", data=df)
```



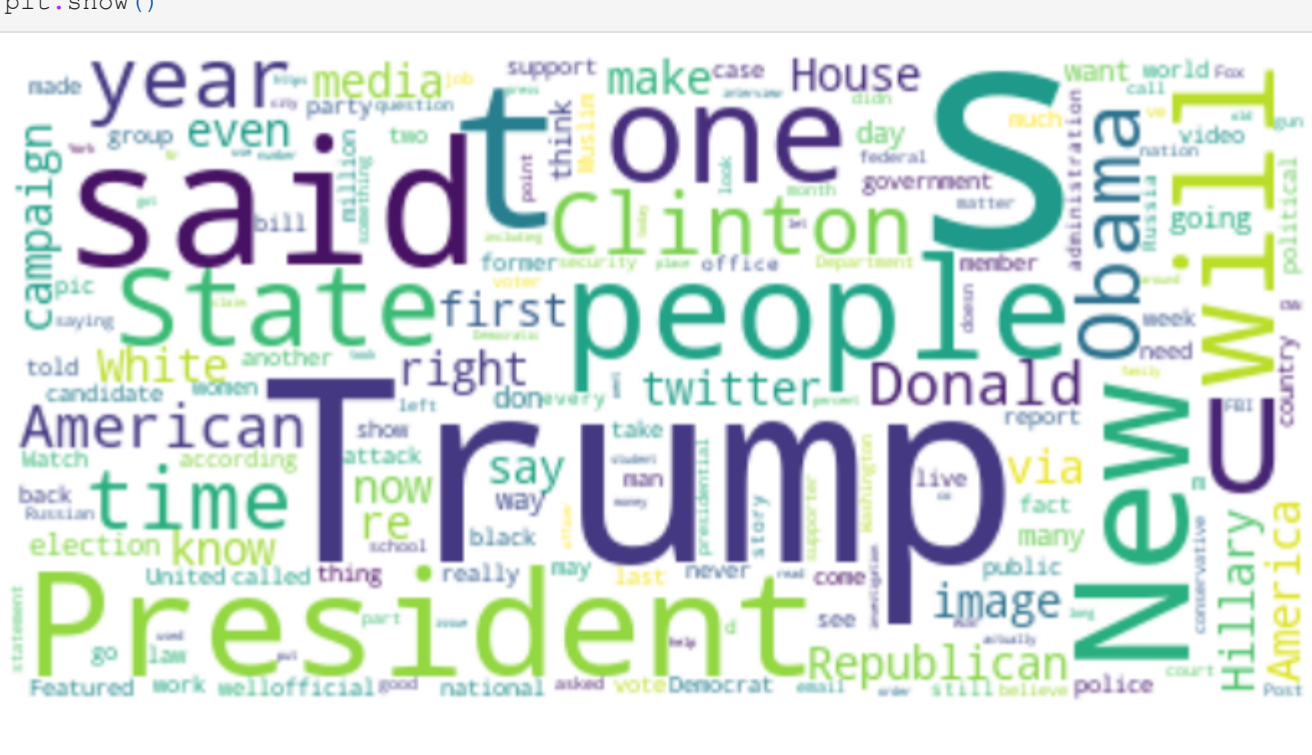
In [10]:
```
df_true = df.loc[df['target'] == 1]
df_fake = df.loc[df['target'] == 0]
```

In [11]:
```
word_cloud_true = WordCloud(collocations = False, background_color = 'white').generate(" ".join(df_true["text"]
word_cloud_fake = WordCloud(collocations = False, background_color = 'white').generate(" ".join(df_fake["text"]
```

In [12]:
```
plt.imshow(word_cloud_true, interpolation='bilinear')
plt.axis("off")
plt.show()
```



In [13]:
```
plt.imshow(word_cloud_fake, interpolation='bilinear')
plt.axis("off")
plt.show()
```



Lots of mentions of Obama and Clinton in the fake news that are not as present in real news. potentially discussing conspiracy theories of misinformation.

In [14]:
```
df["AllText"] = df['subject'] + " " +  df['title'] + " " + df["text"]
```

In [15]:
```
df
```

Out[15]:

	title	text	subject	date	target	AllText
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	1	politicsNews As U.S. budget fight looms, Repub...
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	1	politicsNews U.S. military to accept transgend...
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	1	politicsNews Senior U.S. Republican senator: '...
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	1	politicsNews FBI Russia probe helped by Austra...
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	1	politicsNews Trump wants Postal Service to cha...
...	...	...	...	...	...	...
44893	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0	Middle-east McPain: John McCain Furious That I...
44894	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0	Middle-east JUSTICE? Yahoo Settles E-mail Priv...
44895	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0	Middle-east Sunnistan: US and Allied 'Safe Zon...
44896	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0	Middle-east How to Blow $700 Million: Al Jazee...
44897	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0	Middle-east 10 U.S. Navy Sailors Held by Irani...

44898 rows × 6 columns

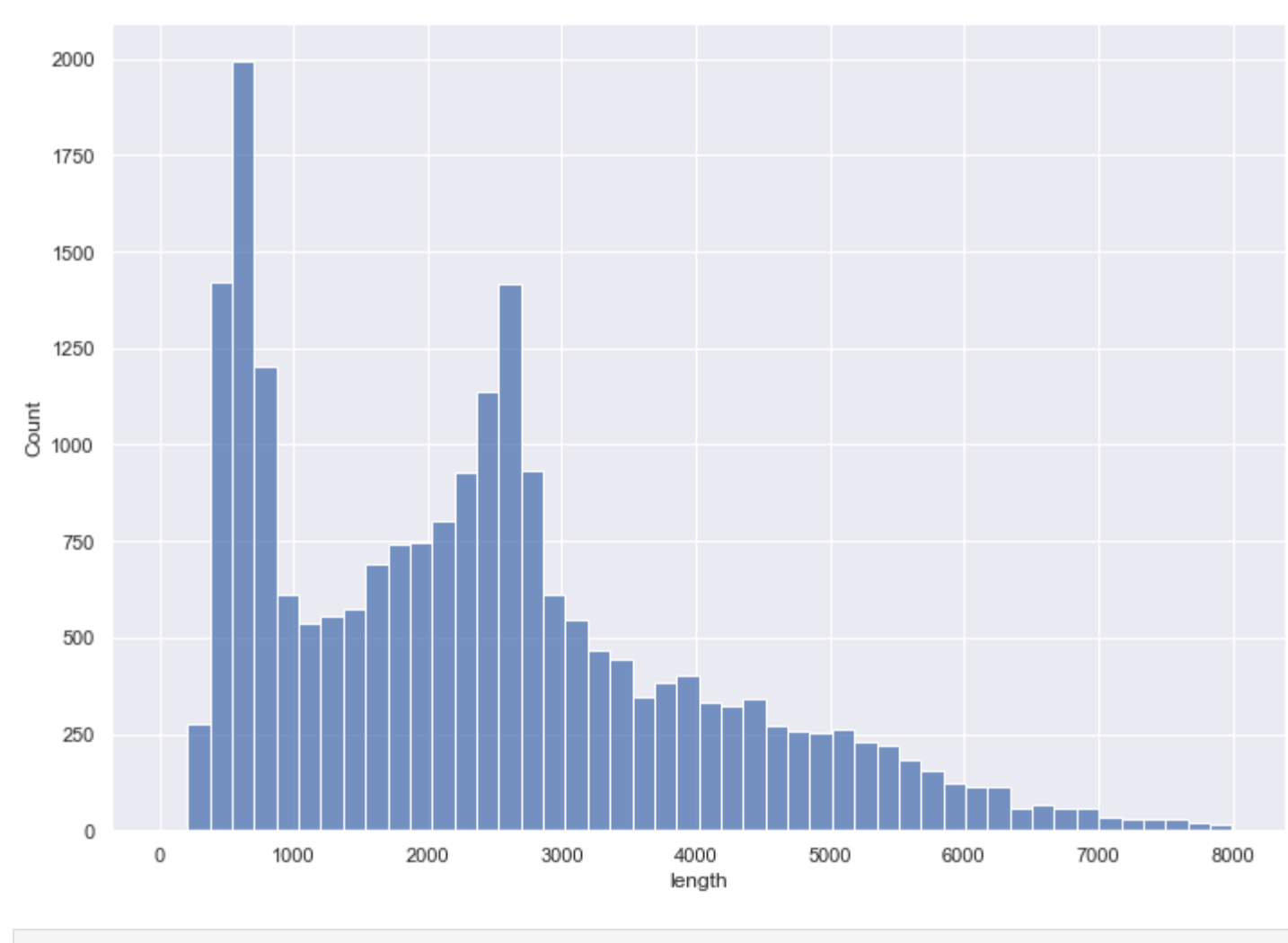
In [16]:
```
df_copy = df.copy()
```

In [17]:
```
df = df.drop(columns = ["title", "text", "subject","date"])
```

In [18]:
```
df["length"] = df['AllText'].apply(len)
```

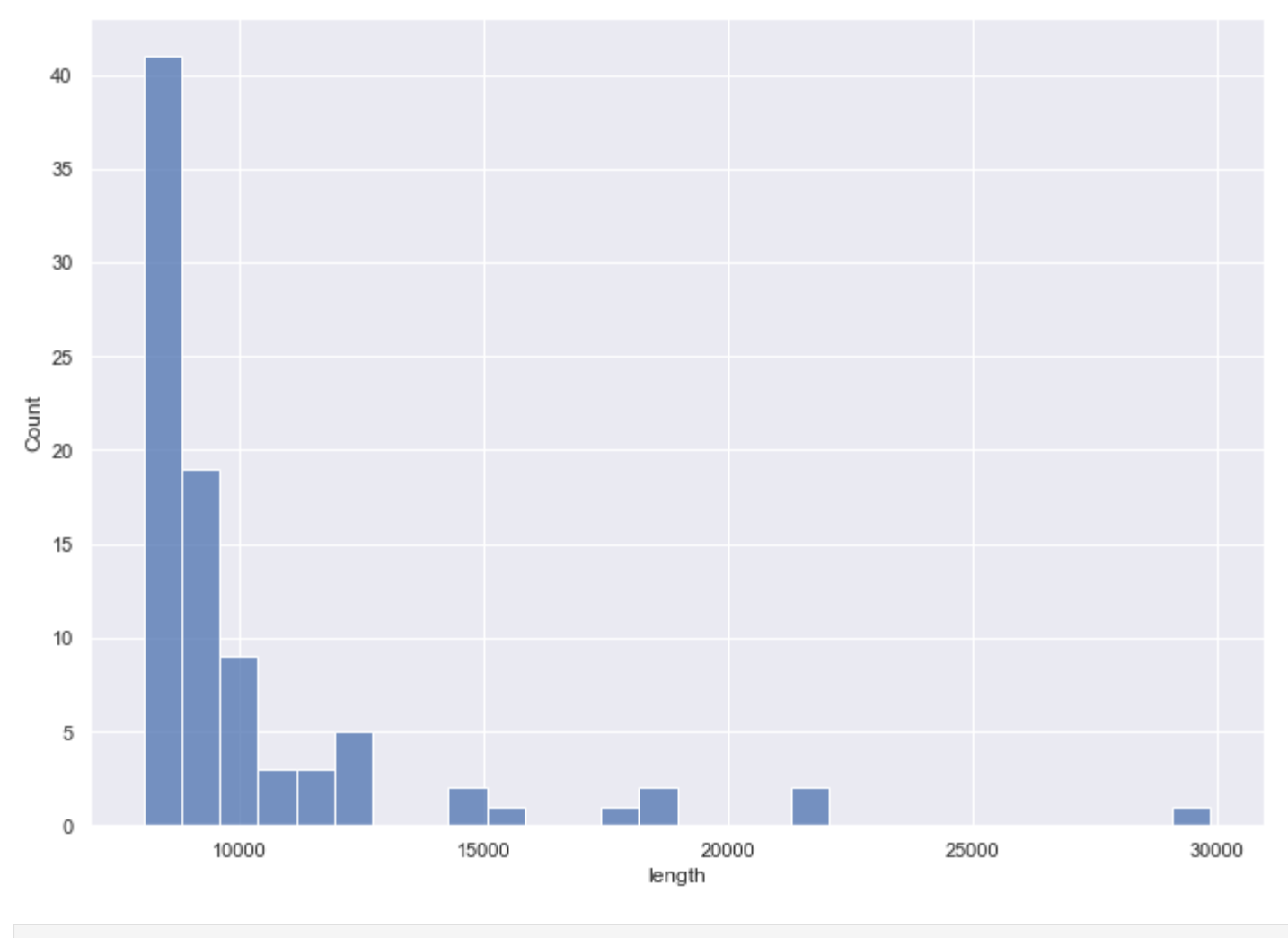
In [19]:
```
sns.histplot(data=df.loc[(df["target"] == 1) & (df["length"] < 8000)], x="length")
```

Out[19]: <AxesSubplot:xlabel='length', ylabel='Count'>



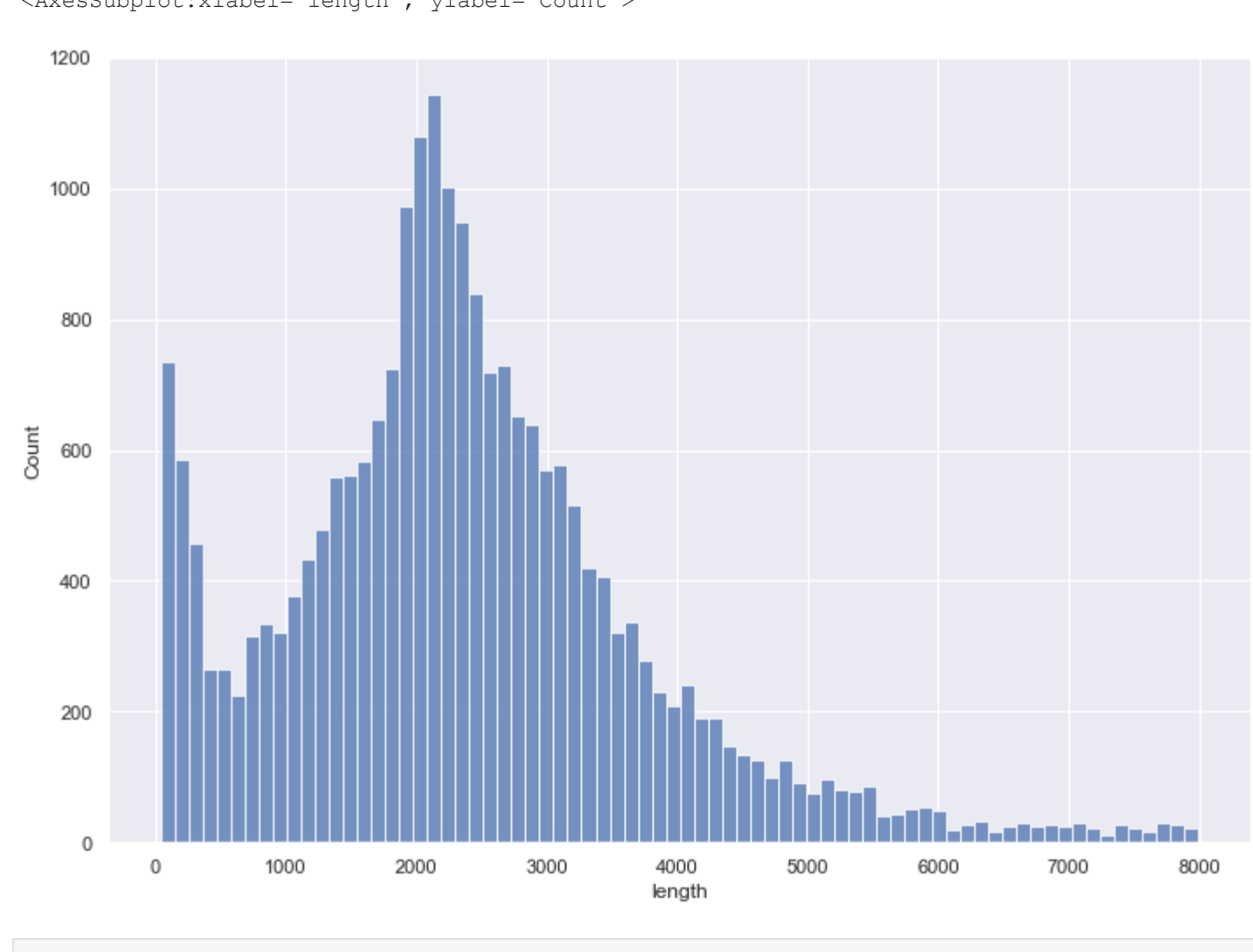
In [20]:
```
sns.histplot(data=df.loc[(df["target"] == 1) & (df["length"] >= 8000)], x="length")
```

Out[20]: <AxesSubplot:xlabel='length', ylabel='Count'>



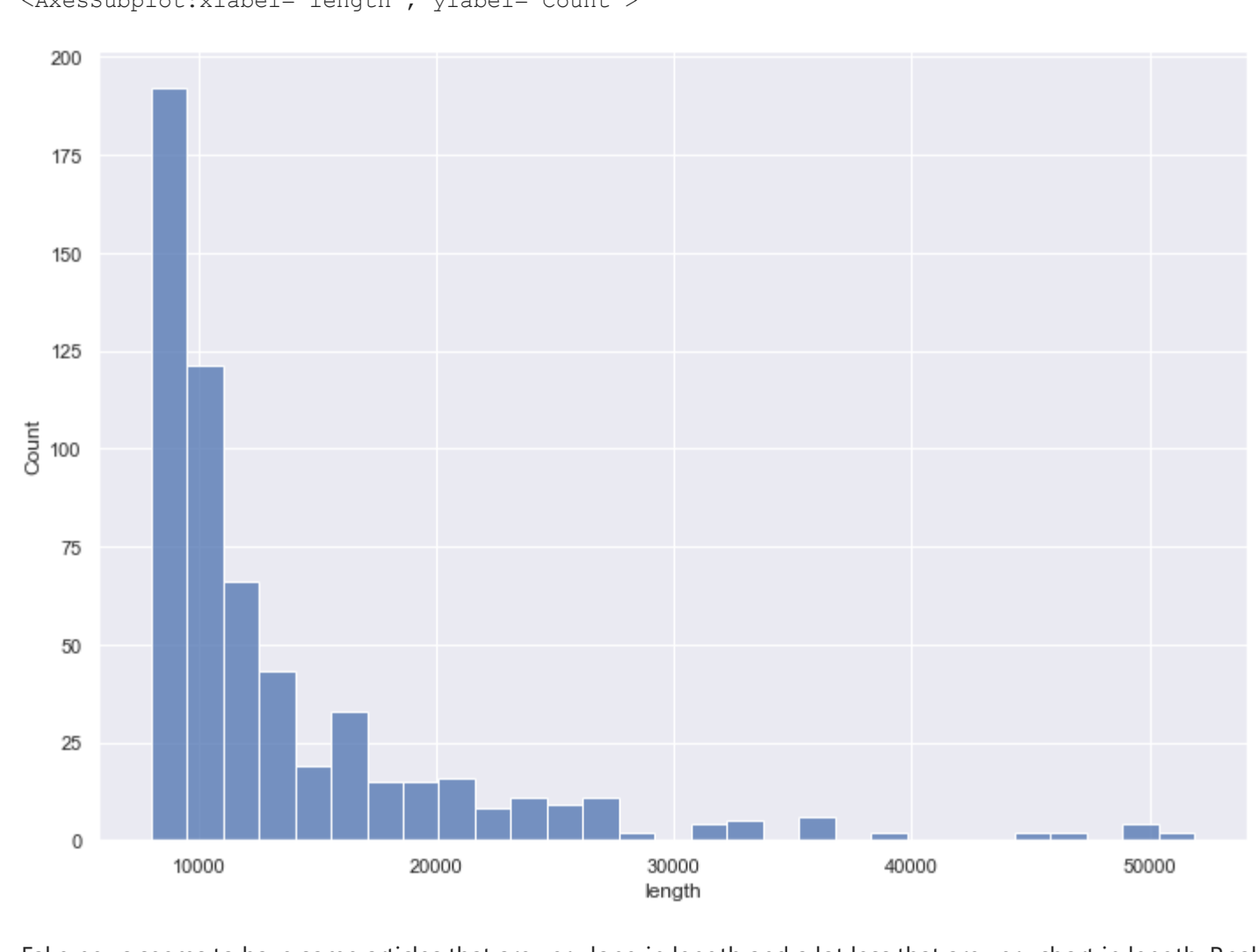
In [21]:
```
sns.histplot(data=df.loc[(df["target"] == 0) & (df["length"] < 8000)], x="length")
```

Out[21]: <AxesSubplot:xlabel='length', ylabel='Count'>



In [22]:
```
sns.histplot(data=df.loc[(df["target"] == 0) & (df["length"] >= 8000)], x="length")
```

Out[22]: <AxesSubplot:xlabel='length', ylabel='Count'>



Fake news seems to have some articles that are very long in length and a lot less that are very short in length. Real news barely has any articles past 8000 words but there are substantial news articles longer than 8000.

TFIDF is a statistic reflecting the importance of a word to a text. It rewards words that are high frequency in few texts while removing weight from words that are common in all texts in order to find words that are most important to the text. Each word from the bag of words can be counted and stored in a count vector which the TFIDF Model can pass over to assign weights.

2. Train Test Split

1. Import related libraries and perform train test split. Keep 20% data in the test set
2. Using a count plot show how many real and fake news do we have in the training set and how many in the test set

In [23]:
```
df
```

Out[23]:

	target	AllText	length
0	1	politicsNews As U.S. budget fight looms, Repub...	4737
1	1	politicsNews U.S. military to accept transgend...	4155
2	1	politicsNews Senior U.S. Republican senator: '...	2863
3	1	politicsNews FBI Russia probe helped by Austra...	2534
4	1	politicsNews Trump wants Postal Service to cha...	5287
...	...	...	...
44893	0	Middle-east McPain: John McCain Furious That I...	3311
44894	0	Middle-east JUSTICE? Yahoo Settles E-mail Priv...	1778
44895	0	Middle-east Sunnistan: US and Allied 'Safe Zon...	25163
44896	0	Middle-east How to Blow $700 Million: Al Jazee...	2765
44897	0	Middle-east 10 U.S. Navy Sailors Held by Irani...	5345

44898 rows × 3 columns

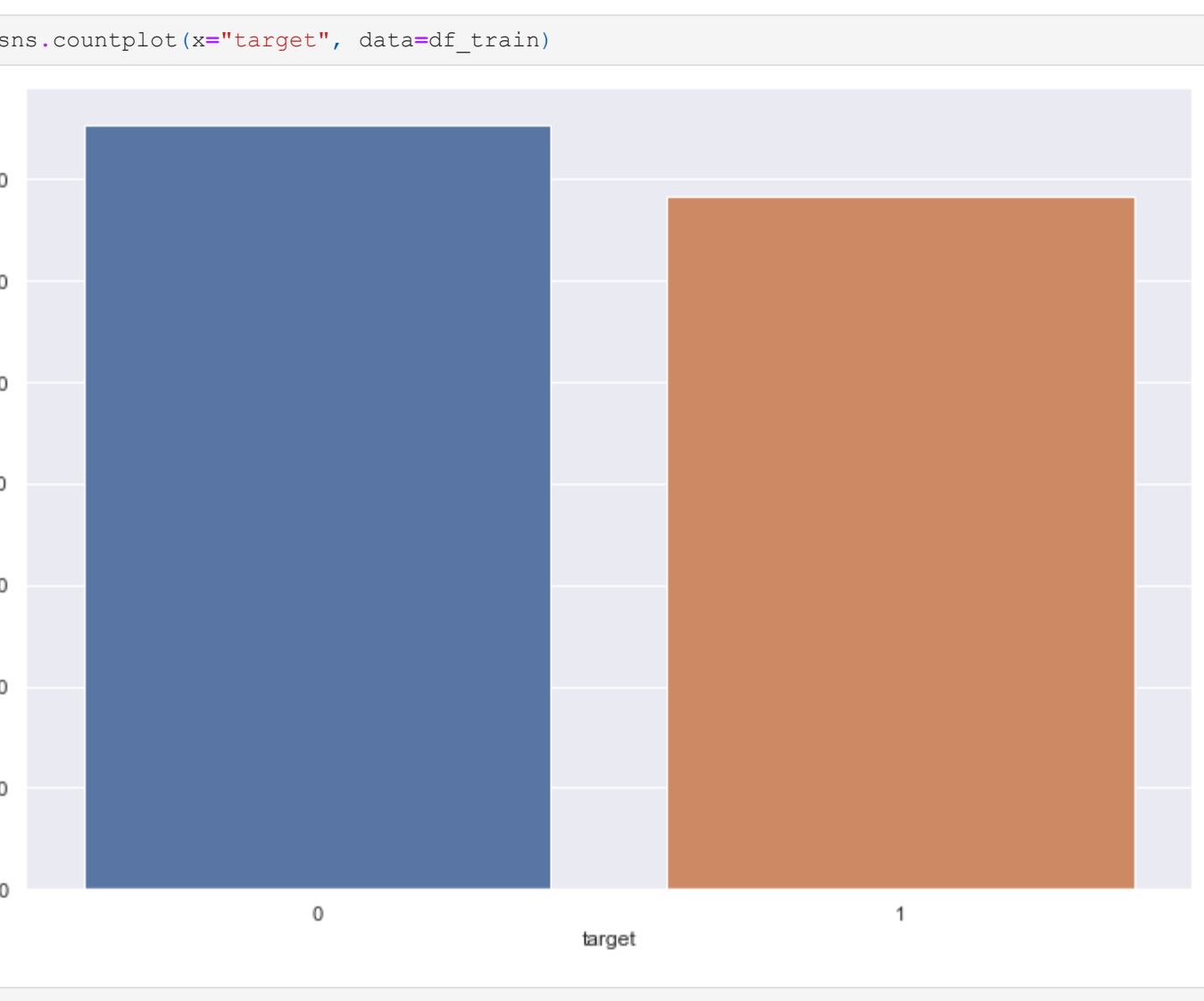
In [24]:
```
X = df.iloc[:, 1:]
y = df.iloc[:, 0]
```

In [25]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=101)
```

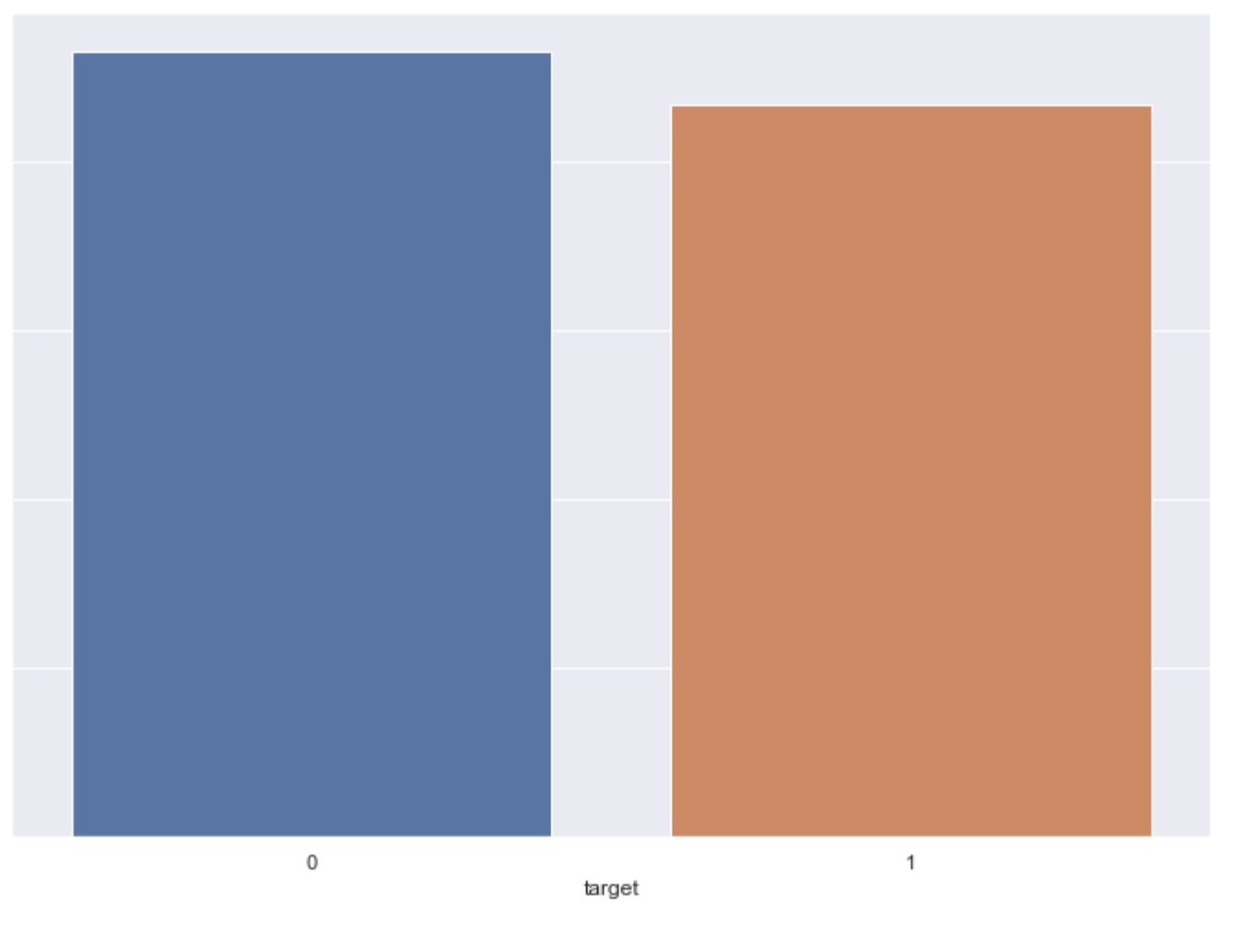
In [26]:
```
df_train = pd.DataFrame()
df_train["AllText"] = X_train["AllText"]
df_train["length"] = X_train["length"]
df_train["target"] = y_train
```

In [27]:
```
df_test = pd.DataFrame()
df_test["AllText"] = X_test["AllText"]
df_test["length"] = X_test["length"]
df_test["target"] = y_test
```

In [28]:
```
ax = sns.countplot(x="target", data=df_train)
```



In [29]:
```
ax = sns.countplot(x="target", data=df_test)
```



3. Training and Testing Fake news classifier using MultinomialNB

1. Create a pipeline that will use countVectorizer with the function you have created earlier for data preprocessing, then use Tftransformer and then use the NaiveBayes classifier
2. Fit the pipeline and then perform prediction
3. Generate classification report and confusion matrix (you have to achieve at least 96% accuracy for the test set to receive full credit)
4. Discuss the performance like how good the model is overall, how good is it in predicting fake news, and how good is it in predicting true news.
5. Copy a part of any news of your choice from a news website and then use the model to predict whether is it true or not.

In [30]:

```python
stemmer = LancasterStemmer()

def text_process(mess):
    """
    Takes in a string of text, then performs the following:
    1. Remove all punctuation
    2. Tokenize
    3. convert them to lower case
    4. Remove all stopwords
    3. Perform stemming
    4. Returns a list of the cleaned text
    """

    stopwordList = stopwords.words('english')

    # Check characters to see if they are in punctuation
    mess = [char for char in mess if char not in string.punctuation]
     # Join the characters again to form the string.
    mess = ''.join(mess)

    words = nltk.word_tokenize(mess)
    words = [t for t in words if t not in stopwordList]
    words = [stemmer.stem(w.lower()) for w in words]

    return words
```

In [31]:

```python
# Check to make sure its working. Just applying the functin to top 5 rows without chaning my data frame
df['AllText'].head(5).apply(text_process)
```

Out[31]:

```
0    [politicsnew, as, us, budget, fight, loom, rep...
1    [politicsnew, us, milit, acceiv, transgend, re...
2    [politicsnew, seny, us, republ, sen, let, mr, ...
3    [politicsnew, fbi, russ, prob, help, aust, dip...
4    [politicsnew, trump, want, post, serv, charg, ...
Name: AllText, dtype: object
```

In [32]:

```python
bow_transformer = CountVectorizer(analyzer=text_process).fit(df['AllText']) #see we are passing our function na

# Print total number of vocab words
print(len(bow_transformer.vocabulary_))
```

```
185013
```

In [33]:

```python
text_bow = bow_transformer.transform(df['AllText'])
text_bow
```

Out[33]:

```
<44898x185013 sparse matrix of type '<class 'numpy.int64'>'
	with 7388188 stored elements in Compressed Sparse Row format>
```

In [34]:

```python
tfidf_transformer = TfidfTransformer().fit(text_bow)
text_tfidf = tfidf_transformer.transform(text_bow)
print(text_tfidf.shape)
```

```
(44898, 185013)
```

In [35]:

```python
text_tfidf
```

Out[35]:

```
<44898x185013 sparse matrix of type '<class 'numpy.float64'>'
	with 7388188 stored elements in Compressed Sparse Row format>
```

In [36]:

```python
pipeline = Pipeline([
    ('bow', CountVectorizer(analyzer=text_process)),  # strings to token integer counts
    ('tfidf', TfidfTransformer()),  # integer counts to weighted TF-IDF scores
    ('classifier', MultinomialNB()),  # train on TF-IDF vectors w/ Naive Bayes classifier
])
```

In [37]:

```python
pipeline.fit(df_train["AllText"],df_train["target"])
```

Out[37]:

```
Pipeline(steps=[('bow',
                 CountVectorizer(analyzer=<function text_process at 0x0000017D08044310>)),
                ('tfidf', TfidfTransformer()),
                ('classifier', MultinomialNB())])
```

In [38]:

```python
predictions = pipeline.predict(df_test["AllText"])
```

In [39]:

```python
print(confusion_matrix(predictions, df_test["target"]))
print(classification_report(predictions,df_test["target"]))
```

```
[[4453  195]
 [ 192 4140]]
              precision    recall  f1-score   support

           0       0.96      0.96      0.96      4648
           1       0.96      0.96      0.96      4332

    accuracy                           0.96      8980
   macro avg       0.96      0.96      0.96      8980
weighted avg       0.96      0.96      0.96      8980
```

Model classifies 96% of articles correctly. There are 195 instances of fake news being classified as real news and 192 instances of real news being classified as fake news.

Fake News

In [40]:

```python
url = "https://babylonbee.com/news/local-carpenter-continues-to-spread-disinformation-and-teachings-deemed-harm
article = Article(url)
```

In [41]:

```python
article.download()
article.parse()
```

In [42]:

```python
text = "Middle-east " + article.title + " " + article.text
```

In [43]:

```python
pipeline.predict([text])
```

Out[43]:

```
array([0], dtype=int64)
```

Missclassified Real News

In [44]:

```python
url = "https://www.foxnews.com/politics/tulsi-gabbard-obama-ministry-truth"
article = Article(url)
```

In [45]:

```python
article.download()
article.parse()
```

In [46]:

```python
text = "US_News " + article.title + " " + article.text
```

In [47]:

```python
pipeline.predict([text])
```

Out[47]:

```
array([0], dtype=int64)
```

Real News

In [48]:

```python
url = "https://www.cnn.com/2022/05/01/europe/russia-farm-vehicles-ukraine-disabled-melitopol-intl/index.html"
article = Article(url)
```

In [49]:

```python
article.download()
article.parse()
```

In [50]:

```python
text = "worldnews " + article.title + " " + article.text
```

In [51]:

```python
pipeline.predict([text])
```

Out[51]:

```
array([1], dtype=int64)
```

4. Training and Testing a Deep Neural Network

1. Import related library for using MLPClassifier from sklearn neural netowrk.
2. Create a pipeline like 3i, for MLPClassfier you should use at least two layers and also should verbose = 2 (you can use other parameters as you wish or use the one you see from the uploaded google colab)
3. Fit the pipeline and then perform prediction
4. Generate classification report and confusion matrix (You have to achieve at least 99% accuracy to receive full credit for this model)
5. Discuss the performance like how good the model is overall, how good is it in predicting fake news, and how good is it in predicting true news.
6. Use the same news you have used above and then use the model to predict whether is it true or not.
7. Discuss any difference in performance between this model and NB model

In [52]:

```python
pipeline = Pipeline([
    ('bow', CountVectorizer(analyzer=text_process)),  # strings to token integer counts
    ('tfidf', TfidfTransformer()),  # integer counts to weighted TF-IDF scores
    ('classifier', MLPClassifier(hidden_layer_sizes=(100,4), random_state=0, early_stopping=True, verbose=2)),
])
```

In [53]:

```python
pipeline.fit(df_train["AllText"],df_train["target"])
```

```
Iteration 1, loss = 0.32756278
Validation score: 0.988864
Iteration 2, loss = 0.18242194
Validation score: 0.991370
Iteration 3, loss = 0.15289204
Validation score: 0.991091
Iteration 4, loss = 0.13295309
Validation score: 0.988864
Iteration 5, loss = 0.11702874
Validation score: 0.988029
Iteration 6, loss = 0.10382493
Validation score: 0.989978
Iteration 7, loss = 0.09263313
Validation score: 0.989143
Iteration 8, loss = 0.08301853
Validation score: 0.990256
Iteration 9, loss = 0.07469006
Validation score: 0.990535
Iteration 10, loss = 0.06743518
Validation score: 0.991370
Iteration 11, loss = 0.06107118
Validation score: 0.990535
Iteration 12, loss = 0.05547103
Validation score: 0.990535
Iteration 13, loss = 0.05050487
Validation score: 0.991370
Validation score did not improve more than tol=0.000100 for 10 consecutive epochs. Stopping.
```

Out[53]:

```
Pipeline(steps=[('bow',
                 CountVectorizer(analyzer=<function text_process at 0x0000017D08044310>)),
                ('tfidf', TfidfTransformer()),
                ('classifier',
                 MLPClassifier(early_stopping=True, hidden_layer_sizes=(100, 4),
                               random_state=0, verbose=2))])
```

In [54]:

```python
predictions = pipeline.predict(df_test["AllText"])
```

In [55]:

```python
print(confusion_matrix(predictions, df_test["target"]))
print(classification_report(predictions,df_test["target"]))
```

```
[[4626   55]
 [  19 4280]]
              precision    recall  f1-score   support

           0       1.00      0.99      0.99      4681
           1       0.99      1.00      0.99      4299

    accuracy                           0.99      8980
   macro avg       0.99      0.99      0.99      8980
weighted avg       0.99      0.99      0.99      8980
```

Model is 99% accurate. 55 Instances of fake news are classifed as real news and 19 instances of real news are classified as fake news

Fake News

In [56]:

```python
url = "https://babylonbee.com/news/local-carpenter-continues-to-spread-disinformation-and-teachings-deemed-harm
article = Article(url)
```

In [57]:

```python
article.download()
article.parse()
```

In [58]:

```python
text = "Middle-east " + article.title + " " + article.text
```

In [59]:

```python
pipeline.predict([text])
```

Out[59]:

```
array([0], dtype=int64)
```

Misclassified Real News

In [60]:

```python
url = "https://www.foxnews.com/politics/tulsi-gabbard-obama-ministry-truth"
article = Article(url)
```

In [61]:

```python
article.download()
article.parse()
```

In [62]:

```python
text = "US_News " + article.title + " " + article.text
```

In [63]:

```python
pipeline.predict([text])
```

Out[63]:

```
array([0], dtype=int64)
```

Misclassified Real News

In [64]:

```python
url = "https://www.cnn.com/2022/05/01/europe/russia-farm-vehicles-ukraine-disabled-melitopol-intl/index.html"
article = Article(url)
```

In [65]:

```python
article.download()
article.parse()
```

In [66]:

```python
text = "worldnews " + article.title + " " + article.text
```

In [67]:

```python
pipeline.predict([text])
```

Out[67]:

```
array([0], dtype=int64)
```

Overall the Deep Learning model is more accurate based on the classification report. This is due to it being a more complex model and thus being able to learn more about the intricacies of the data. It is interesting though on using articles from today, that both the models missclassified real news. This could be due to the topics covered in todays news being different and the model is not reflecting those changes, thus the classifier shows them as fake news. This issue could just be fixed with a newer dataset.

5. Extra Credit

Build a high-performance model for classifying the news into different subjects. It means your target column will be the subject column. During this process, you have to use a neural network, properly preprocess the data frame, remove irrelevant columns, train and test the model properly and then finally show the classification report and confusion matrix. In order to get full credit, it should be at least 79% accurate in predicting the subject of news. Finally, you should discuss the classification report like what are the weaknesses of the model and what are the strengths of the model.

In [ ]:

```python
def data():

    stemmer = LancasterStemmer()

    def text_process(mess):


        stopwordList = stopwords.words('english')

        mess = [char for char in mess if char not in string.punctuation]
        mess = ''.join(mess)

        words = nltk.word_tokenize(mess)
        words = [t for t in words if t not in stopwordList]
        words = [stemmer.stem(w.lower()) for w in words]

        return words

    def convert_sparse_matrix_to_sparse_tensor(X):
        coo = X.tocoo()
        indices = np.mat([coo.row, coo.col]).transpose()
        return tf.SparseTensor(indices, coo.data, coo.shape)

    df = pd.read_csv("news.csv", sep = "\t")
    df["target"] = df["subject"]
    df["TitleText"] = df["title"] + " " + df["text"]
    df = df.drop(columns = ["title", "text", "subject", "date"])
    df.target = df["target"].astype("category").cat.codes
    subject = tf.keras.utils.to_categorical(df["target"].values, num_classes=8)
    text_train, text_test, y_train, y_test = train_test_split(df["TitleText"], subject, test_size=0.2, random_s
    bow_transformer = CountVectorizer(analyzer=text_process).fit(df["TitleText"])
    text_bow = bow_transformer.transform(df["TitleText"])
    tfidf_transformer = TfidfTransformer().fit(text_bow)
    train_bow = bow_transformer.transform(text_train)
    train_tfidf = tfidf_transformer.transform(train_bow)
    test_bow = bow_transformer.transform(text_test)
    test_tfidf = tfidf_transformer.transform(test_bow)
    train_tensor = convert_sparse_matrix_to_sparse_tensor(train_tfidf)
    x_train = tf.sparse.reorder(train_tensor)
    test_tensor = convert_sparse_matrix_to_sparse_tensor(test_tfidf)
    x_test = tf.sparse.reorder(test_tensor)
    return x_train, y_train, x_test, y_test
```

In [13]:

```python
x_train, y_train, x_test, y_test = data()
```

In [176…

```python
model = tf.keras.Sequential()
model.add(layers.Dense(12, input_dim = x_train.shape[1], activation = "tanh")) # input layer requires input_dim
model.add(layers.Dense(12, activation = "tanh"))
model.add(layers.Dense(12, activation = "tanh"))
model.add(layers.Dense(12, activation = "tanh"))
model.add(layers.Dense(8, activation='softmax'))
model.summary()
model.compile(loss="binary_crossentropy", optimizer= "rmsprop", metrics=['accuracy'])
es = tf.keras.callbacks.EarlyStopping(monitor='loss', min_delta=0.005, patience=1, verbose=1, mode='auto')
with tf.device('/cpu:0'):
    model.fit(x_train, y_train, epochs = 100, shuffle = True, batch_size=128, verbose=1, callbacks=[es])
    score = model.evaluate(x_test, y_test, verbose=0)
```

```
Model: "sequential_91"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_461 (Dense)           (None, 12)                2220120

 dense_462 (Dense)           (None, 12)                156

 dense_463 (Dense)           (None, 12)                156

 dense_464 (Dense)           (None, 12)                156

 dense_465 (Dense)           (None, 8)                 104

=================================================================
Total params: 2,220,692
Trainable params: 2,220,692
Non-trainable params: 0
_________________________________________________________________
Epoch 1/100
C:\Users\matth\anaconda3\envs\ML\lib\site-packages\tensorflow\python\framework\indexed_slices.py:444: UserWarning: Converting sparse IndexedSlices(IndexedSlices(indices=Tensor("gradient_tape/sequential_91/dense_461/embedding_lookup_sparse/Reshape_1:0", shape=(None,), dtype=int32), values=Tensor("gradient_tape/sequential_91/dense_461/embedding_lookup_sparse/Reshape:0", shape=(None, 12), dtype=float32), dense_shape=Tensor("gradient_tape/sequential_91/dense_461/embedding_lookup_sparse/Cast:0", shape=(2,), dtype=int32))) to a dense Tensor of unknown shape. This may consume a large amount of memory.
  warnings.warn(
281/281 [==============================] - 2s 6ms/step - loss: 0.3670 - accuracy: 0.3105
Epoch 2/100
281/281 [==============================] - 2s 6ms/step - loss: 0.2288 - accuracy: 0.7060
Epoch 3/100
281/281 [==============================] - 2s 6ms/step - loss: 0.1414 - accuracy: 0.8087
Epoch 4/100
281/281 [==============================] - 2s 7ms/step - loss: 0.0996 - accuracy: 0.8276
Epoch 5/100
281/281 [==============================] - 2s 7ms/step - loss: 0.0826 - accuracy: 0.8408
Epoch 6/100
281/281 [==============================] - 2s 6ms/step - loss: 0.0749 - accuracy: 0.8444
Epoch 7/100
281/281 [==============================] - 2s 7ms/step - loss: 0.0703 - accuracy: 0.8453
Epoch 7: early stopping
```

In [177…

```python
predictions = model.predict(x_test)
```

In [178…

```python
y_predictions=np.argmax(predictions, axis=1)
tests=np.argmax(y_test, axis=1)
```

In [179…

```python
print(confusion_matrix(y_predictions, tests))
print(classification_report(y_predictions,tests))
```

```
[[   1    1    1    1    0    2    0    0]
 [   0    2    0    4    0    0    0    0]
 [   1    2 1744    2   11   30    3    0]
 [   3  162    1  146    0    0    0    0]
 [   0    0    0    0    1    6    0    0]
 [ 280    0   53    0  887 1280    4    9]
 [   1    0    2    0    0    5 2285   27]
 [   4    1    3    1    3    4   40 1967]]
              precision    recall  f1-score   support

           0       0.00      0.17      0.01         6
           1       0.01      0.33      0.02         6
           2       0.97      0.97      0.97      1793
           3       0.95      0.47      0.63       312
           4       0.00      0.14      0.00         7
           5       0.96      0.51      0.67      2513
           6       0.98      0.98      0.98      2320
           7       0.98      0.97      0.98      2023

    accuracy                           0.83      8980
   macro avg       0.61      0.57      0.53      8980
weighted avg       0.97      0.83      0.88      8980
```

Model is 83% accurate in predicting the subject of news. It has weaknesses where support is low. Which is for subjects where there are not many articles for. This simply shows that given more data on articles that have less represented subjects, the model can be improved significantly.

In [ ]: