```
from google.colab import drive
drive.mount('/content/drive')
import nltk
import pandas as pd
import numpy as np
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_rem
```

```
#get data
df = pd.read_csv('/content/drive/MyDrive/Horizon/fake_or_real_news.csv')
df = df.drop('title', axis = 1)
df = df.drop('Unnamed: 0', axis = 1)
```

```
df
```

	text	label
0	Daniel Greenfield, a Shillman Journalism Fello...	FAKE
1	Google Pinterest Digg Linkedin Reddit Stumbleu...	FAKE
2	U.S. Secretary of State John F. Kerry said Mon...	REAL
3	— Kaydee King (@KaydeeKing) November 9, 2016 T...	FAKE
4	It's primary day in New York and front-runners...	REAL
...	...	...
6330	The State Department told the Republican Natio...	REAL
6331	The 'P' in PBS Should Stand for 'Plutocratic' ...	FAKE
6332	Anti-Trump Protesters Are Tools of the Oligar...	FAKE
6333	ADDIS ABABA, Ethiopia —President Obama convene...	REAL
6334	Jeb Bush Is Suddenly Attacking Trump. Here's W...	REAL

6335 rows × 2 columns

```
fake_news = df["text"][df['label']=="FAKE"].values
real_news = df["text"][df['label']=="REAL"].values
```

```
#only keep text data in a list
#fake_news = fake_df['text'].values
#real_news = real_df['text'].values
```

```
len(fake_news[34])
```

```
2687
```

```
print(len(fake_news))
print('------------')
print(len(real_news))
```

```
3164
------------
3171
```

Pre-process raw text for sentiment analysis

```
processed_fake_news = []
processed_real_news = []
import re
print(type(processed_real_news))
type(fake_news[43])
```

```
<class 'list'>
str
```

```
#remove hyperlinks
for i in range(len(fake_news)):
  t = re.sub(r'https\S+', '', fake_news[i])
  processed_fake_news.append(t)
for i in range(len(real_news)):
  t = re.sub(r'https\S+', '', real_news[i])
  processed_real_news.append(t)
len(processed_fake_news)
```

```
3164
```

```
print(fake_news[87])
print("----------")
print(processed_fake_news[87])
```

Only Making Matters Worse in Syria October 28, 2016
Exclusive: Washington's foreign policy establishment is determined to escalate U.S. military attacks in Syria even †
By Daniel Lazare
Middle East policy has reached an inflexion point, a moment when Official Washington seems to be caught in the midd
On one hand, the rhetoric has not been more militant since Hillary Clinton's famous "we came, we saw, he died" momen
Thus, The New York Times assures us that that the Middle East is "desperate for American leadership" while the Wash:
Leading think tanks are publishing "a flurry of reports" urging stepped-up intervention, including U.S.-backed "safe
As Vox puts it: "The hot new policy idea in Washington is the hottest old idea: direct US military intervention in S
But reading between the lines, a very different picture emerges, a realization that the U.S. has painted itself into
While reporting that pro-escalation sentiment is unanimous in Washington's vast foreign-policy establishment — somet
"My concern is that we may be talking to each other and agreeing with each other," it quotes one expert as saying, "
Official Washington in a Bubble
Thus, even the Establishment worries that it lives in a bubble. Washington wants war, it needs war, and yet it admit
Then there are the mild liberals over at Vox, the hip and successful Washington website founded by journalistic wund
As Beauchamp lectures away amid fancy graphics and cool background music, the video faithfully toes the Washington '
But the real turning point, he says, occurred in September 2015 when Russia stepped in with airstrikes that allowed
"A siege," Beauchamp then explains, "involves trapping a group of people, civilians and fighters both, inside a cert
The upshot, he says, is "a humanitarian crisis … roughly 250,000 people trapped in the city … running dangerously lc
"The United States has the military power to break the siege of Aleppo," Beauchamp says, "but doing so would be extr
So U.S. options are zero: "Every diplomatic solution tried so far has failed, and failed miserably, and there's simp
Falsehoods and Obfuscations
Beauchamp's explanation — which is really not an explanation at all, merely an assertion — is studded with falsehooc
While asserting that 250,000 people are trapped inside east Aleppo, he ignores reports that the real figure is far '
Beauchamp also implies that it is the government that has "trapped" people in east Aleppo when reports in the Londor
When Salafists launched a short-lived offensive in east Aleppo last summer, for instance, The New York Times reporte
One could go on, but the point is clear. The foreign-policy establishment has not only cut itself off from the publi
So how will the United States respond now that Washington is preparing for a transition between Obama-style abstenti
Washington will continue to fume as the catastrophe deepens in both Syria and Iraq. As the drive to push ISIS (also
The fall of east Aleppo, which now seems to be a foregone conclusion, will drive the foreign-policy establishment tc
With that, Zack Beauchamp will star in another Vox video entitled, "U.S. military intervention in Syria, explained :
When Russian soldiers are killed, he'll try to shift blame to Vladimir Putin for stirring up trouble in the eastern
While anything can go wrong with this scenario, one thing is clear. The mood in Official Washington is the opposite
Daniel Lazare is the author of several books including The Frozen Republic: How the Constitution Is Paralyzing Democ
----

```
#remove other links
for i in range(len(processed_fake_news)-1):
  processed_fake_news[i] = re.sub(r'pic.twitter.com/[\w]*', '', processed_fake_news[i])
for i in range(len(processed_real_news)-1):
  processed_real_news[i] = re.sub(r'pic.twitter.com/[\w]*', '', processed_real_news[i])
```

```
#print(fake_news[23480])
#print("-----------")
#print(processed_fake_news[23480])
```

Convert to lowercase

```
for i in range(len(processed_fake_news)):
  processed_fake_news[i] = processed_fake_news[i].lower()
for i in range(len(processed_real_news)):
  processed_real_news[i] = processed_real_news[i].lower()
```

```
print(real_news[2340])
print("-----------")
print(processed_real_news[2340])
```

```
In a Monday column for the far-right website World Net Daily, the longtime anti-feminist crusader lamented the declin

"Long ago when I went to college, campuses were about 70 percent male, and until 1970 it was still nearly 60 percent,

Never one to shirk victim-blaming, Schlafly proceeded to link the problem of campus sexual assault to the increased e

"Boys are more likely than girls to look at the cost-benefit tradeoff of going to college," Schlafly asserted. "The i

With so many women around, what do you expect a college man to do — seek consent!? "So, what's the solution?" Schlafl
-----------
in a monday column for the far-right website world net daily, the longtime anti-feminist crusader lamented the declin

"long ago when i went to college, campuses were about 70 percent male, and until 1970 it was still nearly 60 percent,

never one to shirk victim-blaming, schlafly proceeded to link the problem of campus sexual assault to the increased e

"boys are more likely than girls to look at the cost-benefit tradeoff of going to college," schlafly asserted. "the i

with so many women around, what do you expect a college man to do — seek consent!? "so, what's the solution?" schlafl
```

Tokenize the articles

```
from nltk.tokenize import word_tokenize
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
True
```

```
for i in range(len(processed_fake_news)):
  processed_fake_news[i] = word_tokenize(processed_fake_news[i])
for i in range(len(processed_real_news)):
  processed_real_news[i] = word_tokenize(processed_real_news[i])
```

```
print(fake_news[2348])
print("-----------")
print(processed_fake_news[2348])
```

```
LIFESTYLE
Finland's education system is considered one of the best in the world. In international ratings, it's always in the t
Finnish officials want to remove school subjects from the curriculum. There will no longer be any classes in physics,
The head of the Department of Education in Helsinki, Marjo Kyllonen, explained the changes:
"There are schools that are teaching in the old-fashioned way which was of benefit in the beginning of the 1900s — bu
Instead of individual subjects, students will study events and phenomena in an interdisciplinary format. For example,
This system will be introduced for senior students, beginning at the age of 16. The general idea is that the students
The traditional format of teacher-pupil communication is also going to change. Students will no longer sit behind sch
The Finnish education system encourages collective work, which is why the changes will also affect teachers. The scho
The changes are expected to be complete by 2020.
What do you think about all these ideas? We'd love to hear your opinion, so let us know in the comments.
Preview photo credit ZouZou
-----------
['lifestyle', 'finland', ''', 's', 'education', 'system', 'is', 'considered', 'one', 'of', 'the', 'best', 'in', 'the'
```

Get rid of stopwords and punctuation

```
nltk.download('stopwords')
from nltk.corpus import stopwords      #remove stopwords
import string    #remove punctuation
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]    Package stopwords is already up-to-date!
```

```
stopwords_english = stopwords.words('english')
```

```
for i in range(len(processed_fake_news)):
  t = processed_fake_news[i]
  q = []
  for w in t:
    if (w not in stopwords_english and w not in string.punctuation):
      q.append(w)
    processed_fake_news[i] = q
for i in range(len(processed_real_news)):
  q = []
  t = processed_real_news[i]
  for w in t:
    if (w not in stopwords_english and w not in string.punctuation):
      q.append(w)
    processed_real_news[i] = q
```

```
print(fake_news[234])
print("-----------")
print(processed_fake_news[234])
```

```
Unprecedented Surge In Election Fraud Incidents From Around The Country    Zero Hedge
Mounting evidence would suggest it's getting more and more difficult for the left to claim that there are "no signs"
A 74-year-old woman tasked with opening envelopes sent by Miami-Dade County voters with their completed mail ballots

Investigators linked Gladys Coego, a temporary worker for the county elections department, to two fraudulent votes, b

In a separate election-fraud case, authorities also arrested a second woman for unlawfully filling out voter-registra

The Miami-Dade state attorney's office plans to accuse Tomika Curgil, 33, of filling out forms for five people withou

Police officers arrested Curgil at her Liberty City home Friday morning and intend to charge her with five felony cou
Miami-Dade State Attorney Katherine Fernández Rundle was quick to praise the "swift arrest of the wrongdoers" and ens
"Our law enforcement effort against these election law violators was swift and resulted in an immediate arrest of the

"Anyone who attempts to undermine the democratic process should recognize that there is an enforcement partnership be

"I want to ensure the voters of Miami-Dade County that the integrity of the electoral process is intact because our p
Meanwhile, Florida isn't the only state with fraud problems as an NBC affiliate in Virginia is reporting that a forme
A former resident of Alexandria, Virginia, is facing up to 40 years in prison after he allegedly used fake names to f

Vafalay Massaquoi, 30, is facing four felony charges related to allegations of voter registration fraud , the Commonw

In the spring of 2016, Massaquoi was registering new voters as an employee of a local advocacy group. According to th

The fake applications were filed with the Alexandria Office of the General Registrar, who reported the issue to Commo
All of these reports simply add fuel to the fire of Trump who has been relentlessly attacking the "rigged" elections
— Donald J. Trump (@realDonaldTrump) October 16, 2016 'Obama Warned Of Rigged Elections In 2008.' Time to #DrainTheSw
— Donald J. Trump (@realDonaldTrump) October 25, 2016

Of course, these are just a few of the people who have actually been caught for their election "rigging" efforts. Mea
Susie Sallee was buried in 1998. Yet records show she voted in Chicago 12 years later.

Victor Crosswell died in 1994, but records show he's voted six times since then.

And then there's Floyd Stevens. Records show he's voted 11 times since his death in 1993.

"It's crazy," Sharon Stevens Anderson, Stevens' daughter, tells CBS 2's Pam Zekman. "I don't see how people can be ab

Those are just a few of the cases CBS 2 Investigators found by merging Chicago Board of Election voter histories with

In all, the analysis showed 119 dead people have voted a total of 229 times in Chicago in the last decade.

Moreover, an ABC affiliate in Philadelphia uncovered similar instances of dead voters in the "City of Brotherly Love.
So, Action News dug through a decade's worth of election and death records to see if there was any truth to the claim

Some of what Action News investigation found was stunning.

Pezzano passed in 2006 . But state voting records show the South Philadelphia native still listed as an "Active Voter

Our investigation also found Joseph B. Haggarty resting peacefully in a Bucks County cemetery. His grave marker confi

Action News also found Paul Bunch, who died in 2006, also cast a vote in this year's primary which was nearly ten yea
```

```
    But, while all of this may seem shocking, in due time, we're confident these arrests and all other instances dead peo
    -----------
    ['unprecedented', 'surge', 'election', 'fraud', 'incidents', 'around', 'country', 'zero', 'hedge', 'mounting', 'evide
```

Stemming

```
from nltk.stem import PorterStemmer
stemmer = PorterStemmer()
```

```
for i in range(len(processed_fake_news)):
  t = processed_fake_news[i]
  q = []
  for w in t:
    s = stemmer.stem(w)
    q.append(s)
  processed_fake_news[i] = q
for i in range(len(processed_real_news)):
  t = processed_real_news[i]
  q = []
  for w in t:
    s = stemmer.stem(w)
    q.append(s)
  processed_real_news[i] = q
```

```
print(real_news[2340])
print("-----------")
print(processed_real_news[2340])
```

```
In a Monday column for the far-right website World Net Daily, the longtime anti-feminist crusader lamented the declin

"Long ago when I went to college, campuses were about 70 percent male, and until 1970 it was still nearly 60 percent,

Never one to shirk victim-blaming, Schlafly proceeded to link the problem of campus sexual assault to the increased e

"Boys are more likely than girls to look at the cost-benefit tradeoff of going to college," Schlafly asserted. "The i

With so many women around, what do you expect a college man to do — seek consent!? "So, what's the solution?" Schlafl
-----------
['monday', 'column', 'far-right', 'websit', 'world', 'net', 'daili', 'longtim', 'anti-feminist', 'crusad', 'lament',
```

Add to one list

```
news = processed_real_news + processed_fake_news
```

Create labels

```
import numpy as np
labels = np.append(np.ones(len(processed_real_news)), np.zeros(len(processed_fake_news)))
```

```
len(labels)
```

```
6335
```

Using dictionaries

Finding frequencies of each word for fake or real news

```
#function to find frequency
def fre(news, labels):
  freqs = {}
  for i in range(len(news)):
    t = news[i]
    for w in t:
      pair = (w, labels[i])
      if pair in freqs:
        freqs[pair] += 1
      else:
        freqs[pair] = 1
  return freqs
```

```
#function to return frequency of word given string and fake or real
def freq_word_sen(freq,w,s): # w is a string, s is 0 or 1
  if (w,s) in freq:
    return freq[(w,s)]
  return 0
```

```
freqs = fre(news, labels)
```

```
freq_word_sen(freqs,'tel',0)
```

```
30
```

```
y = labels
```

Create features

frequency of real news is first

frequency of fake news is second

```
def make_features(news,freqs):
  X = np.zeros((len(news),2))        #initialize X to 0 for all features
  for i in range(len(news)):
    t = news[i]
    for w in t:
      X[i,0] = X[i,0] + freq_word_sen(freqs,w,1)     #add frequencies to the postive
      X[i,1] = X[i,1] + freq_word_sen(freqs,w,0)     #add frequencies to the negative
  return X
```

```
X = make_features(news, freqs)
```

```
X
```
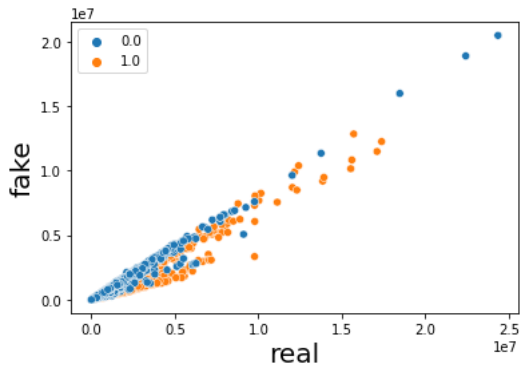
```
array([[ 714825.,   481159.],
       [ 754484.,   297116.],
       [ 133604.,   117272.],
       ...,
       [ 988427.,   787002.],
       [4448491., 3767340.],
       [1555429., 1127145.]])
```

Start training model

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn import preprocessing
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Activation
from tensorflow.keras.losses import BinaryCrossentropy
from sklearn.metrics import classification_report
```

```
sns.scatterplot(X[:,0], X[:,1], hue =y)
plt.xlabel('real',fontsize=20)
plt.ylabel('fake',fontsize=20)
#plt.xlim(0, 1000)
#plt.ylim(0,1000)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43: FutureWarning: Pass the following variables as keyw
  FutureWarning
Text(0, 0.5, 'fake')
```



```python
#split into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size = 0.25, random_state = 4)
```

```python
print(X_train.shape)
print(X_val.shape)
print(y_train.shape)
print(y_val.shape)
```

```
(4751, 2)
(1584, 2)
(4751,)
(1584,)
```

```python
#scaling the features
scaler = preprocessing.StandardScaler()
scaler.fit(X_train)
X_train_scaled = scaler.transform(X_train)
X_val_scaled = scaler.transform(X_val)
```

```python
#function for decision boundaries
def decision_boundary():
  xl, xr, dx = 0, 1000, 10
  yl, yr, dy = 0, 1000, 10
  u = np.arange(xl,xr,dx)
  v = np.arange(yl,yr,dy)
  u_r = np.ones((len(v),1))*u.reshape(1,len(u))
  v_r = v.reshape(len(v),1)*np.ones((1,len(u)))
  u_r = u_r.reshape(-1)
  v_r = v_r.reshape(-1)
  p_grid = np.column_stack((u_r,v_r))
  p_grid_scaled = scaler.transform(p_grid)
  f_grid = model.predict(p_grid_scaled)
  f_grid = f_grid.reshape((len(v),len(u)))
  plt.contour(u,v,f_grid,levels=[0.5])
  return
```
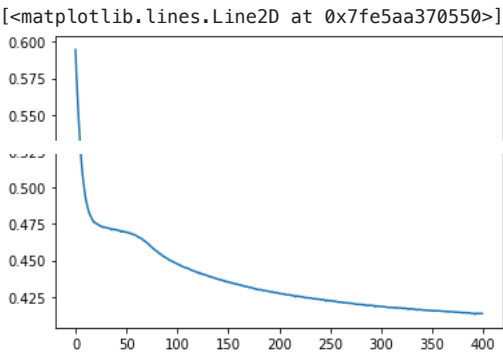
```
Start coding or generate with AI.
```

```python
# computing error
bce = BinaryCrossentropy(from_logits= False)
```

```python
J_val = []
J_train = []
```

Creating model

```python
model = 0
model = Sequential()     #first layer
model.add(Dense(4, activation = "tanh"))
model.add(Dense(1, activation= "sigmoid"))     #output bc binary classification
model.compile(loss = 'binary_crossentropy')  #error to minimize
model.fit(X_train_scaled, y_train, epochs = 400, verbose = 0)   #create model
```

```
J_list = model.history.history['loss']
plt.plot(J list)
```

```
[<matplotlib.lines.Line2D at 0x7fe5aa370550>]
```



```
y_train_hat = model.predict(X_train_scaled)
J_train.append(bce(y_train.reshape(-1,1), y_train_hat).numpy())
print(J_train)
print("-------")
y_val_hat = model.predict(X_val_scaled)
J_val.append(bce(y_val.reshape(-1,1), y_val_hat).numpy())
print(J_val)
```

```
[0.48293242, 0.6931261, 0.50283146, 0.496712, 0.44842097, 0.4125897, 0.41097635, 0.41076314, 0.4117996, 0.40919012, 0
-------
[0.45843667, 0.6932544, 0.48885095, 0.48168737, 0.4506996, 0.4073719, 0.4068618, 0.4069195, 0.41001222, 0.41016373, 0
```

```
#sns.scatterplot(X[:,0],X[:,1],hue=y)
#decision_boundary()
#plt.xlabel('feature 1',fontsize=20)
#plt.ylabel('feature 2',fontsize=20)
#plt.xlim(0,1e7)
#plt.ylim(0,1e7)
```

```
y_train_hat_cat = 1*(model.predict(X_train_scaled) > 0.5)
y_val_hat_cat = 1*(model.predict(X_val_scaled)> 0.5)
print(classification_report(y_train,y_train_hat_cat))
print(classification_report(y_val, y_val_hat_cat))
```

	precision	recall	f1-score	support
0.0	0.78	0.86	0.82	2360
1.0	0.85	0.76	0.80	2391
accuracy			0.81	4751
macro avg	0.82	0.81	0.81	4751
weighted avg	0.82	0.81	0.81	4751

	precision	recall	f1-score	support
0.0	0.79	0.85	0.82	804
1.0	0.83	0.77	0.80	780
accuracy			0.81	1584
macro avg	0.81	0.81	0.81	1584
weighted avg	0.81	0.81	0.81	1584

Start coding or generate with AI.