

PHASE 2 : FAKENEWS DETECTION USING NLP

Introduction :

Fake news detection using NLP is the process of using natural language processing techniques to identify and filter out false or misleading information within textual content , such as news articles , to promote accurate information dissemination.

Employ semantic analysis techniques to understand the meaning of text beyond individual words. This can help in detecting fake news by identifying inconsistencies, contradictions, or unusual patterns in the text that might not be apparent with traditional keyword-based methods.

Be aware of adversarial attacks where malicious actors attempt to circumvent fake news detection system. Regularly update and fortify your NLP models to withstand these attacks, which may involve intentionally crafted misinformation.

In [1] :

```
import numpy as np
import pandas as pd
import plotly.express as px
import plotly.graph_objs as go
from plotly.subplots import make_subplots

import nltk
from nltk.corpus import stopwords
import tensorflow as tf
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ModelCheckpoint
from sklearn.model_selection import train_test_split
from transformers import AutoTokenizer, TFAutoModelForSequenceClassification

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
nltk.download('stopwords')
```

Out[1]:

True

In [2]:

```
fake_news_path = "/kaggle/input/fake-and-real-news-dataset/Fake.csv"
real_news_path = "/kaggle/input/fake-and-real-news-dataset/True.csv"
```

In [3]:

```
fake_news = pd.read_csv(fake_news_path)
real_news = pd.read_csv(real_news_path)
```

In [4]:

```
fake_news.head(3)
```

Out[4]:

	Title	Text	subject	Date
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017

In [5]:

```
real_news.head(3)
```

Out[5]:

	Title	Text	subject	Date
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017

In [6]:

```
real = real_news.copy()
```

```
fake = fake_news.copy()
```

In [7]:

```
real['Label'] = 'Real'  
fake['Label'] = 'Fake'
```

In [8]:

```
news = pd.concat([real, fake], axis=0, ignore_index=True)  
news.reset_index()  
news.head()
```

Out[8]:

	Title	text	subject	date	Label
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	Real
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	Real
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	Real
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	Real
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	Real

In [9]:

```
print(f"Samples available: {news.shape[0]}\n#features of dataset: {news.shape[1]}")
```

Samples available: 44898

#features of dataset: 5

In [10]:

```
news_ds = news.sample(1000).drop(['title', 'date', 'subject'], axis=1)  
news_ds.head(3)
```

Out[10]:

	Text	Label
36923	If local law enforcement begins to act like i...	Fake
38038	Obama and HUD want to give one last freebie to...	Fake
28271	Corey Lewandowski got some good news last week...	Fake

In [11]:

```
CLASS_NAMES = ['Fake', 'Real']
class_mapper = {
    'Fake':0,
    'Real':1
}
```

In [12]:

```
news_ds['Label'] = news_ds['Label'].map(class_mapper)
news_ds.head(3)
```

Out[12]:

	Text	Label
36923	If local law enforcement begins to act like i...	0
38038	Obama and HUD want to give one last freebie to...	0
28271	Corey Lewandowski got some good news last week...	0

In [13]:

```
class_dist = px.histogram(data_frame=news,
                           y='Label',
                           color='Label',
                           title='Fake vs Real news Original dataset',
                           text_auto=True)
class_dist.update_layout(showlegend=False)
```

```
class_dist.show()
21.417k23.481k05k10k15k20kFakeReal
Fake vs Real news Original datasetcountLabel
```

In [14]:

```
subject_dist = px.histogram(data_frame=news,
                             x='subject',
                             color='subject',
                             title='Fake vs Real news Subject Distribution',
                             text_auto=True,
                             facet_col='Label')
subject_dist.update_layout(showlegend=False)
subject_dist.show()
11.272k10.145kpoliticsNewsworldnewsNewsPoliticsGovernment Newsleft-newsUS_NewsMiddle-
east02k4k6k8k10k9050684115704459783778politicsNewsworldnewsNewsPoliticsGovernment Newsleft-
newsUS_NewsMiddle-east
Fake vs Real news Subject DistributionsubjectsubjectcountLabel=RealLabel=Fake
```

In [15]:

```
news.date.unique().max()
```

Out[15]:

```
'https://fedup.wpengine.com/wp-content/uploads/2015/04/hillarystreetart.jpg'
```

In [16]:

```
list(filter(lambda x:len(x)>20, news.date.unique()))
```

Out[16]:

```
['https://100percentfedup.com/served-roy-moore-vietnamletter-veteran-sets-record-st-
raight-honorable-decent-respectable-patriotic-commander-soldier/',
 'https://100percentfedup.com/video-hillary-asked-about-trump-i-just-want-to-eat-so-
me-pie/',
 'https://100percentfedup.com/12-yr-old-black-conservative-whose-video-to-obama-wen-
t-viral-do-you-really-love-america-receives-death-threats-from-left/',
 'https://fedup.wpengine.com/wp-content/uploads/2015/04/hillarystreetart.jpg',
 'https://fedup.wpengine.com/wp-content/uploads/2015/04/entitled.jpg',
 'MSNBC HOST Rudely Assumes Steel Worker Would Never Let His Son Follow in His Foot-
steps...He Couldn't Be More Wrong [Video]']
```

In [17]:

```
news = news[news['date'].map(lambda x:len(x)) <= 20]
news.date = pd.to_datetime(news['date'], format='mixed')
news.head()
```

Out[17]:

	Title	Text	subject	date	Label
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	2017-12-31	Real
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	2017-12-29	Real
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	2017-12-31	Real
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	2017-12-30	Real
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	2017-12-29	Real

In [18]:

```
date_dist = px.histogram(data_frame=news,
                          x='date',
                          color='Label')
date_dist.show()
Jul 2015Jan 2016Jul 2016Jan 2017Jul 2017Jan 20180100200300400500600700800900
LabelRealFakedatecount
```

In [19]:

```
subject_dist = px.histogram(data_frame=news,
                             x='date',
                             color='subject')
subject_dist.show()
2016201720180100200300400500600700800900
subjectpoliticsNewsworldnewsNews politicsGovernment Newsleft-newsUS_NewsMiddle-eastdatecount
```

In [20]:

```
real_sub_dist = px.histogram(data_frame=news[news['Label']=='Real'],
                              x='date',
                              color='subject')
real_sub_dist.show()
Jul 2016Jan 2017Jul 2017Jan 20180200400600800100012001400
subjectpoliticsNewsworldnewsdatecount
```

Text Preprocessing :

In [21]:

```
import string
```

In [22]:

```
stop_words = stopwords.words('english')
def text_preprocessing(text):
    words = text.lower().split()
    filtered_words = [word for word in words if word not in stop_words]
    pure_text = ' '.join(filtered_words)
    pure_text = pure_text.translate(str.maketrans('', '', string.punctuation)).strip()
    return pure_text
```

In [23]:

```
X = news_ds.text.apply(text_preprocessing).to_numpy()
y = news_ds.Label.to_numpy().astype('float32').reshape(-1, 1)

train_X, test_X, train_y, test_y = train_test_split(X, y,
                                                    train_size=0.9,
                                                    stratify=y,
                                                    random_state=7)
train_X, val_X, train_y, val_y = train_test_split(train_X, train_y,
                                                    train_size=0.9,
                                                    stratify=train_y,
                                                    random_state=7)
```

In [24]:

```
model_name = "BERTFakeNewsDetector"
model_callbacks = ModelCheckpoint(model_name, save_best_only=True)
```

In [25]:

```
bert_name = 'bert-base-uncased'
tokenizer = AutoTokenizer.from_pretrained(bert_name,
                                         padding='max_length',
                                         do_lower_case=True,
                                         add_special_tokens=True)
```

Downloading (...)okenizer_config.json: 100%

28.0/28.0 [00:00<00:00, 2.34kB/s]

Downloading (...)lve/main/config.json: 100%

570/570 [00:00<00:00, 44.6kB/s]

Downloading (...)solve/main/vocab.txt: 100%

232k/232k [00:00<00:00, 4.02MB/s]

Downloading (...)main/tokenizer.json: 100%

466k/466k [00:00<00:00, 24.0MB/s]

In [26]:

```
def tokenize(df):
    inputs = tokenizer(df.tolist(),
                      padding=True,
                      truncation=True,
                      return_tensors='tf').input_ids

    return inputs
```

In [27]:

```
train_X_encoded = tokenize(train_X)
val_X_encoded = tokenize(val_X)
test_X_encoded = tokenize(test_X)
```

In [28]:

```
def prepare_datasets(encoded, true_df, true_target_df):
    return tf.data.Dataset.from_tensor_slices((encoded, true_target_df)).shuffle(tr
ue_df.shape[0]).batch(8).prefetch(tf.data.AUTOTUNE)
```

In [29]:

```
train_ds = prepare_datasets(train_X_encoded, train_X, train_y)
test_ds = prepare_datasets(test_X_encoded, test_X, test_y)
val_ds = prepare_datasets(val_X_encoded, val_X, val_y)
```

In [30]:

```
model = TFAutoModelForSequenceClassification.from_pretrained(bert_name,
                                                            num_labels=1)
```

Downloading model.safetensors: 100%

440M/440M [00:02<00:00, 211MB/s]

All PyTorch model weights were used when initializing TFBertForSequenceClassification.

Some weights or buffers of the TF 2.0 model TFBertForSequenceClassification were not initialized from the PyTorch model and are newly initialized: ['classifier.weight', 'classifier.bias']

You should probably TRAIN this model on a down-stream task to be able to use it for predictions and inference.

In [31]:

```
model.compile(
    optimizer = Adam(learning_rate=1e-5),
    metrics = [
        tf.keras.metrics.BinaryAccuracy(name='Accuracy'),
        tf.keras.metrics.Precision(name='Precision'),
        tf.keras.metrics.Recall(name='Recall')
    ]
)
```

```
model_history = model.fit(train_ds,
                          validation_data=val_ds,
                          callbacks=model_callbacks,
                          epochs=5,
                          batch_size=16)
```

```
model_history = pd.DataFrame(model_history.history)
```

Epoch 1/5

102/102 [=====] - 185s 1s/step - loss: 0.1815 - Accuracy: 0.7358 - Precision: 0.7548 - Recall: 0.6337 - val_loss: 0.0347 - val_Accuracy: 0.9667 - val_Precision: 0.9535 - val_Recall: 0.9762

Epoch 2/5

102/102 [=====] - 134s 1s/step - loss: 0.0457 - Accuracy: 0.9444 - Precision: 0.9410 - Recall: 0.9385 - val_loss: 0.0205 - val_Accuracy: 0.9889 - val_Precision: 0.9767 - val_Recall: 1.0000

Epoch 3/5


```

102/102 [=====] - 137s 1s/step - loss: 0.0153 - Accuracy:
0.9926 - Precision: 0.9946 - Recall: 0.9893 - val_loss: 0.0092 - val_Accuracy: 0.98
89 - val_Precision: 0.9767 - val_Recall: 1.0000
Epoch 4/5
102/102 [=====] - 135s 1s/step - loss: 0.0072 - Accuracy:
0.9963 - Precision: 0.9947 - Recall: 0.9973 - val_loss: 0.0064 - val_Accuracy: 0.98
89 - val_Precision: 0.9767 - val_Recall: 1.0000
Epoch 5/5
102/102 [=====] - 136s 1s/step - loss: 0.0045 - Accuracy:
1.0000 - Precision: 1.0000 - Recall: 1.0000 - val_loss: 0.0025 - val_Accuracy: 1.00
00 - val_Precision: 1.0000 - val_Recall: 1.0000

```

In [32]:

```
model_history
```

Out[32]:

	Loss	Accuracy	Precision	Recall	val_loss	val_Accuracy	val_Precision	val_Recall
0	0.181528	0.735802	0.754777	0.633690	0.034657	0.966667	0.953488	0.97619
1	0.045669	0.944444	0.941019	0.938503	0.020463	0.988889	0.976744	1.00000
2	0.015262	0.992593	0.994624	0.989305	0.009228	0.988889	0.976744	1.00000
3	0.007200	0.996296	0.994667	0.997326	0.006398	0.988889	0.976744	1.00000
4	0.004542	1.000000	1.000000	1.000000	0.002476	1.000000	1.000000	1.00000

In [33]:

```
model.save(model_name)
```

In [34]:

```
fig = make_subplots(rows=2, cols=2, subplot_titles=('Loss', 'Accuracy', 'Precision',
, 'Recall'))
```

```
fig.add_trace(go.Scatter(y=model_history['loss'], mode='lines', name='Training Loss'), row=1, col=1)
```

```
fig.add_trace(go.Scatter(y=model_history['val_loss'], mode='lines', name='Validation Loss'), row=1, col=1)
```

```

fig.add_trace(go.Scatter(y=model_history['Accuracy'], mode='lines', name='Training Accuracy'), row=1, col=2)
fig.add_trace(go.Scatter(y=model_history['val_Accuracy'], mode='lines', name='Validation Accuracy'), row=1, col=2)

fig.add_trace(go.Scatter(y=model_history['Precision'], mode='lines', name='Training Precision'), row=2, col=1)
fig.add_trace(go.Scatter(y=model_history['val_Precision'], mode='lines', name='Validation Precision'), row=2, col=1)

fig.add_trace(go.Scatter(y=model_history['Recall'], mode='lines', name='Training Recall'), row=2, col=2)
fig.add_trace(go.Scatter(y=model_history['val_Recall'], mode='lines', name='Validation Recall'), row=2, col=2)

fig.update_layout(title='Model Training History')

fig.update_xaxes(title_text='Epoch', row=1, col=1)
fig.update_xaxes(title_text='Epoch', row=1, col=2)
fig.update_xaxes(title_text='Epoch', row=2, col=1)
fig.update_xaxes(title_text='Epoch', row=2, col=2)

fig.update_yaxes(title_text='Loss', row=1, col=1)
fig.update_yaxes(title_text='Accuracy', row=1, col=2)
fig.update_yaxes(title_text='Precision', row=2, col=1)
fig.update_yaxes(title_text='Recall', row=2, col=2)

fig.show()

```

0123400.050.10.15012340.80.91012340.80.91012340.70.80.91
Training LossValidation LossTraining AccuracyValidation AccuracyTraining PrecisionValidation PrecisionTraining RecallValidation RecallModel Training HistoryEpochEpochEpochEpochLossAccuracyPrecisionRecallLossAccuracyPrecisionRecall

In [35]:

```
test_loss, test_acc, test_precision, test_recall = model.evaluate(test_ds, verbose = 0)
```

```

print(f"Test Loss      : {test_loss}")
print(f"Test Accuracy   : {test_acc}")
print(f"Test Precision    : {test_precision}")
print(f"Test Recall       : {test_recall}")

```

```

Test Loss      : 0.003277710871770978
Test Accuracy   : 1.0
Test Precision  : 1.0
Test Recall     : 1.0

```

In [36]:

```

def make_prediction(text, model=model):
    text = np.array([text])
    inputs = tokenize(text)
    return np.abs(np.round(model.predict(inputs, verbose=1).logits))

```

In [37]:

```

for _ in range(5):
    index = np.random.randint(test_X.shape[0])

```

```

text = test_X[index]
real = test_y[index]
model_pred = make_prediction(text)

print(f"Original Text:\n\n{text}\n\nTrue: {CLASS_NAMES[int(real)]}\t\tPredicted
: {CLASS_NAMES[int(model_pred)]}\n\n{'-'*100}\n")
1/1 [=====] - 3s 3s/step

```

Original Text:

reuters republican candidate backed president donald trump us senate seat alabama expected finish distant second primary election tuesday still enough votes advance runoff local news projected election seat vacated republican jeff sessions became u s attorney general luther strange appointed seat backed trump move republican primary along former alabama supreme court chief justice roy moore projections said neither candidate captured 50 percent vote required outright victory 60 state's 67 counties reporting moore 438 percent vote strange 317 percent official results showed strange former alabama attorney general close ties senate majority leader mitch mcconnell feuding trump lately stalled effort pass health care reform strange scored president's surprise endorsement last week trump reiterated support strange automated phone call voters monday "he's helping senate" trump said "he's going get tax cuts us he's lot things people alabama" moore effectively ousted alabama's chief justice 2016 defying us supreme court's ruling legalizing gay marriage second time suspended bench first 2003 refusing federal court order remove monument ten commandments courthouse us representative mo brooks came third republican primary advance sept 26 runoff top candidates battled weeks supportive trump's legislative agenda democratic primary doug jones former us attorney backed former vice president joe biden projected outright winner long shot win seat republican dominated state

True: Real Predicted: Real

1/1 [=====] - 3s 3s/step

Original Text:

fbi announced clearing democratic nominee hillary clinton recent email probe donald trump responded epic meltdown front everyone freedom hill amphitheater michigan sun day in speaking crowd republican nominee whined moaned election rigged made clear still thinks guilty ever trump shouted right protected rigged system totally rigged system saying long time review 650000 emails eight days folks hillary clinton guilty hillary clinton guilty knows it fbi knows it people know american people deliver justice ballot box 8 november watch trump throw temper tantrum below donald trump responds fbi letter review 650000 new emails 8 days httpstcon3lykbkln7 httpstcof2ad7fs1 41 abc news abc november 7 2016 for record fbi actually cover kind groundwork eight days since donald ever interested thinking logically sharing facts completely missing irony would say next trump warned investigations crimes go long long time trump obviously projecting onto clinton currently thousands lawsuits him including high profile sexual assault cases recently come light lewd comments audio tape uncovered fact paid taxes decades anyone worried guilty wrongdoing definitely trump something keep mind election day tomorrow featured image via joe raedle getty images

True: Fake Predicted: Fake

1/1 [=====] - 0s 54ms/step

Original Text:

washington reuters us president donald trump said interview friday right impose tax imports mexico options could "much positive" countries "it's something right do it's something impose want" trump said interview christian broadcasting network trump noting spoken mexican president enrique pena nieto said two countries "getting along actually well" "so i'm something like respect mexico something else could happen would much positive mexico united states" trump said

True: Real

Predicted: Real

1/1 [=====] - 0s 96ms/step

Original Text:

american citizens eligible apply small business loans also responsible paying back interest illegal aliens hand getting free money american taxpayer start business obligation pay us back oh one big catch read fine print small business administration website find agency provide funds start grow small business engine american economy providing roughly threequarters new jobs rather sba provides offers loan guarantees banks lenders provide seed money might come surprise american entrepreneurs least one group federal government providing direct assistance business startups illegal aliens fact recipients taxpayer funds fall even restricted category illegal aliens deported back native landas part ongoing project calls waste report sen rand paul rk y called attention outlays made obscure federal agency called interamerican foundation iaf housed nondescript building blocks white house iaf spending 50000 annually help salvadorans sent back el salvador reintegrate communities idea make less likely seek return united states iaf works local ngo called insami stands instituto salvadore del migrante estimates 500 salvadorans deported us el salvador week group says iaf funds assist sixty salvadorans year including deportees helping facilitate reintegration communities support enterprises accomplished offering financial education technical advice assistance business plans serves turn assure broader support resources available migrants abilities appreciated concerns understood needs met iaf respond fox news requests comment deport them set business makes sense me sen paul said interview fox newss hawn moran vice president national border patrol council nbpc union represents america 21000 border patrol officers expressed dismay learning federal outlays deported illegal aliens moran indicated limited funding made exceedingly rare border patrol officers ever partner working along southern border often night state affairs leaves officers vulnerable assaults heavily armed alien drug smugglers also alluded equipment shortages compromise officers effectiveness agents tried true tools need conduct law enforcement operations moran told fox news think 50000 could go long way one border patrol sector help correct that cited communications devices officers rely on agents using radios oftentimes see border patrol station hills along border yet talk it else might 50000 taxpayer funds purchase secure border illegal aliens el salvador elsewhere a review help wanted ads posted online federal customs border protection cbp agency finds recently april 7 cbp conducting search border patrol agents multiple locations top annual starting salary placed at 50600 for amount federal government could purchase seven yamaha grizzly 450 automatic utility all terrain vehicles like these federal government purchased fifteen decatur illinois based motor sports store september 2013 via fox news

True: Fake

Predicted: Fake

1/1 [=====] - 0s 100ms/step

Original Text:

detroitnew york reuters us president donald trump's "buy american hire american" executive order tuesday left questions government would enforce order whether would make real difference output employment according steel executives analysts "buy american" provisions already exist us law policing difficult waivers granted foreign companies undercut us counterparts pricing earlier tuesday trump ordered review government procurement rules favoring american companies see actually benefiting especially us steel industry trump's executive order promises properly police provisions avoided detail happen bill hickey president chicagobased laphamhickey steel seven steel mills midwest northeast said heard talk "buy american" decades american foreign contractors frequently find loopholes use imported steel "politicians talk same end day doesn't work" hickey said citing waivers existing provisions charles bradford bradford research said focusing "buy american" us steel take account steel products including tin plate semifinished products made united states enforced improperly could cause supply problems us market 25 percent steel imported first quarter year "the people pushed don't clue don't know math" said bradford cutting supply goods made united states would create fresh problems us companies said construction industry also concerns "too strict definition constitutes usmade steel products" said kenneth simonson chief economist associated general contractors america simonson cited concerns steel might melted scrap metal could come outside united states example tracing origins point trump's white house track record far also helps fuel skepticism inside industry instead bold action promised last year then candidate trump north american free trade agreement china free trade agreements new administration "not shown much evidence so" said keybank capital markets steel analyst philip gibbs "i'm lot less optimistic threeandahalf months ago far i've seen coming trump administration prior administration" added result gibbs said investors dial back expectations trump anything meaningful trade infrastructure order could make difference investors seemed shrug tuesday's executive order nucor corp shares closed 02 percent 5733 ak steel holding corp gained penny end 632 united states steel corp closed 05 percent 2873 ak steel respond requests comment nucor us steel welcomed president's executive order move welcomed labor unions united steelworkers said current practice "contractors often try avoid law loopholes buy cheap often substandard foreign products like many china" thomas gibson chief executive lobby group american iron steel institute said statement "buy american" provisions "are vital health domestic steel industry helped create manufacturing jobs build american infrastructure" veteran steel industry analyst michelle applebaum said remains seen thoroughly trump administration police steel industry executive order sends clear message steel importers "trump created risk anyone wants import steel" said "if puts money behind enforcement force people play rules good thing"

True: Real

Predicted: Real

Conclusion :

Thus the fake news detection using NLP ,the above techniques are involved in the NLP. The codes are available in document and output also.

