

Fake News Detection Using Machine Learning Techniques

```
In [47]: from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn import feature_extraction, linear_model, model_selection, preprocessing
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.utils import shuffle
from sklearn.tree import DecisionTreeClassifier
import pandas as pd
import numpy as np
import pickle
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn import metrics
import itertools
```

Load datasets

```
In [48]: fake_dataset = pd.read_csv(r"Fake_News.csv");
true_dataset = pd.read_csv(r"Real_News.csv");
```

```
In [49]: fake_dataset.shape
```

```
Out[49]: (23481, 4)
```

```
In [50]: true_dataset.shape
```

```
Out[50]: (21417, 4)
```

Dataset cleaning

```
In [51]: fake_dataset['target'] = 'fake'
true_dataset['target'] = 'true'
```

```
In [52]: fake_dataset.head()
```

Out[52]:

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

```
In [53]: true_dataset.head()
```

Out[53]:

	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	true
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	true
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	true
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	true
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	true

```
In [54]: dataset = pd.concat([fake_dataset, true_dataset]).reset_index(drop = True)
dataset.shape
```

Out[54]: (44898, 5)

```
In [55]: dataset.head(5)
```

Out[55]:

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

```
In [56]: dataset.tail(5)
```

Out[56]:

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

In [57]:

```
dataset = shuffle(dataset)
dataset = dataset.reset_index(drop=True)
```

In [58]:

```
dataset.head()
```

Out[58]:

		title	text	subject	date	target
0	EPIC! TUCKER TAKES ON THE TERROR APPEASERS ON ...	Tucker Carlson nails it tonight on how we re n...	politics	May 23, 2017	fake	
1	MARK LEVIN Destroys Joe Scarborough...Defends Tr...	Conservative talk show host Mark Levin shredde...	politics	Jun 30, 2017	fake	
2	House Republicans move to shut down Democratic...	WASHINGTON (Reuters) - U.S. House of Represent...	politicsNews	June 23, 2016	true	
3	HILARIOUS! Look Who LIBERAL Middlebury Profess...	Two weeks ago at Middlebury College, Charles M...	left-news	Mar 14, 2017	fake	
4	WATCH: NJ MAN MAY FACE JAIL For Flying Donald ...	This is just one more fine example of why we c...	politics	Apr 8, 2016	fake	

In [59]:

```
dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44898 entries, 0 to 44897
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   title       44898 non-null  object
1   text        44898 non-null  object
2   subject     44898 non-null  object
3   date        44898 non-null  object
4   target      44898 non-null  object
dtypes: object(5)
memory usage: 1.7+ MB
```

In [60]:

```
dataset.drop(["date"],axis=1,inplace=True)
dataset.head()
```

Out[60]:

	title	text	subject	target
0	EPIC! TUCKER TAKES ON THE TERROR APPEASERS ON ...	Tucker Carlson nails it tonight on how we re n...	politics	fake
1	MARK LEVIN Destroys Joe Scarborough...Defends Tr...	Conservative talk show host Mark Levin shredde...	politics	fake
2	House Republicans move to shut down Democratic...	WASHINGTON (Reuters) - U.S. House of Represent...	politicsNews	true
3	HILARIOUS! Look Who LIBERAL Middlebury Profess...	Two weeks ago at Middlebury College, Charles M...	left-news	fake
4	WATCH: NJ MAN MAY FACE JAIL For Flying Donald ...	This is just one more fine example of why we c...	politics	fake

```
In [61]: dataset.drop(["title"],axis=1,inplace=True)
dataset.head()
```

Out[61]:

	text	subject	target
0	Tucker Carlson nails it tonight on how we re n...	politics	fake
1	Conservative talk show host Mark Levin shredde...	politics	fake
2	WASHINGTON (Reuters) - U.S. House of Represent...	politicsNews	true
3	Two weeks ago at Middlebury College, Charles M...	left-news	fake
4	This is just one more fine example of why we c...	politics	fake

```
In [62]: dataset['text'] = dataset['text'].apply(lambda x: x.lower())
dataset.head()
```

Out[62]:

	text	subject	target
0	tucker carlson nails it tonight on how we re n...	politics	fake
1	conservative talk show host mark levin shredde...	politics	fake
2	washington (reuters) - u.s. house of represent...	politicsNews	true
3	two weeks ago at middlebury college, charles m...	left-news	fake
4	this is just one more fine example of why we c...	politics	fake

```
In [63]: import string

def punctuation_removal(text):
    all_list = [char for char in text if char not in string.punctuation]
    clean_str = ''.join(all_list)
    return clean_str

dataset['text'] = dataset['text'].apply(punctuation_removal)
```

```
In [64]: dataset.head()
```

```
Out[64]:
```

	text	subject	target
0	tucker carlson nails it tonight on how we re n...	politics	fake
1	conservative talk show host mark levin shredde...	politics	fake
2	washington reuters us house of representative...	politicsNews	true
3	two weeks ago at middlebury college charles mu...	left-news	fake
4	this is just one more fine example of why we c...	politics	fake

```
In [65]: dataset.tail()
```

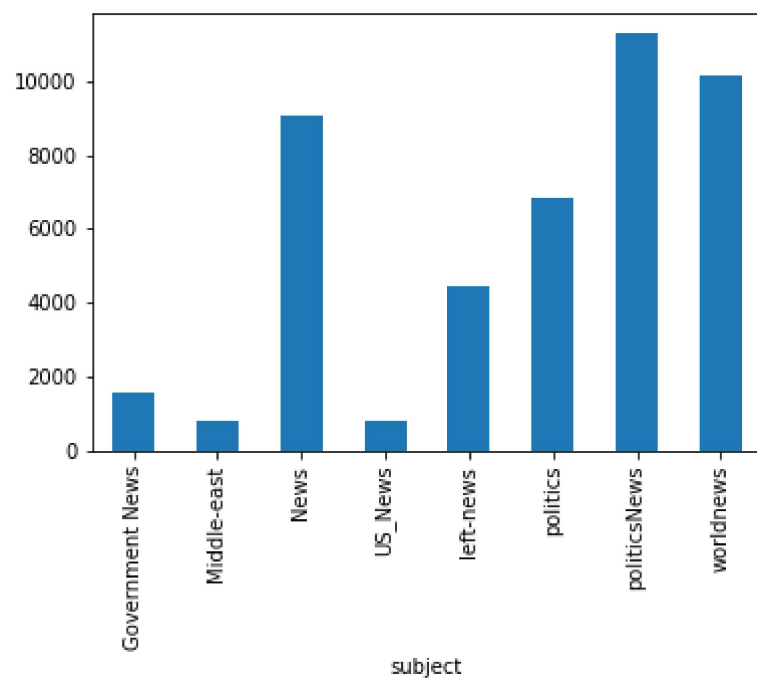
```
Out[65]:
```

	text	subject	target
44893	listen to the mind boggling answers to these q...	left-news	fake
44894	an atlanta democrat makes some very valid poin...	politics	fake
44895	dublin reuters it is not yet clear if britain...	worldnews	true
44896	reuters presidential politics crept into the ...	politicsNews	true
44897	hint it s not because of jane fonda although t...	left-news	fake

Dataset Exploration

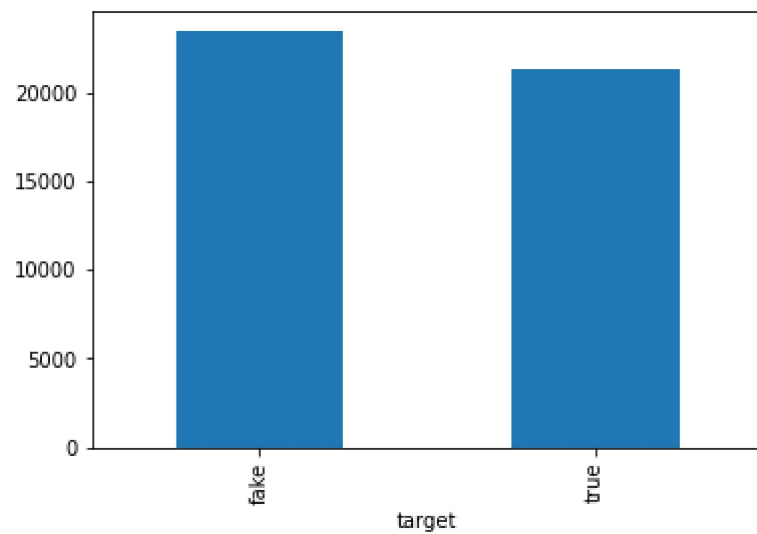
```
In [66]: print(dataset.groupby(['subject'])['text'].count())
dataset.groupby(['subject'])['text'].count().plot(kind="bar")
plt.show()
```

```
subject
Government News    1570
Middle-east        778
News               9050
US_News            783
left-news          4459
politics           6841
politicsNews       11272
worldnews          10145
Name: text, dtype: int64
```



```
In [67]: print(dataset.groupby(['target'])['text'].count())
dataset.groupby(['target'])['text'].count().plot(kind="bar")
plt.show()
```

```
target
fake    23481
true    21417
Name: text, dtype: int64
```



Data Modeling

```
In [68]: def plot_confusion_matrix_from_data(cm, classes,
                                             normalize=False,
                                             title='Confusion matrix',
                                             cmap=plt.cm.Blues):

    plt.imshow(cm, interpolation='nearest', cmap=cmap)
    plt.title(title)
    plt.colorbar()
    tick_marks = np.arange(len(classes))
    plt.xticks(tick_marks, classes, rotation=45)
    plt.yticks(tick_marks, classes)

    if normalize:
        cm = cm.astype('float') / cm.sum(axis=1)[:, np.newaxis]
        print("Normalized confusion matrix")
    else:
        print('Confusion matrix, without normalization')

    thresh = cm.max() / 2.
    for i, j in itertools.product(range(cm.shape[0]), range(cm.shape[1])):
        plt.text(j, i, cm[i, j],
                 horizontalalignment="center",
                 color="white" if cm[i, j] > thresh else "black")

    plt.tight_layout()
    plt.ylabel('True label')
    plt.xlabel('Predicted label')
```

Data Splitting

```
In [69]: X_train, X_test, y_train, y_test = train_test_split(dataset['text'], dataset.target, test_size=0.2, random_state=42)
```

```
In [70]: X_test.head
```

```
Out[70]: <bound method NDFrame.head of 22216 donald trump wanted to celebrate power like a ...
27917 once again the gun grabbers are doing anything...
25007 tune in to the alternate current radio network...
1377 washington reuters us president donald trump ...
32476 washington reuters the white house on tuesday...
...
42119 the function of socialism is to raise sufferin...
4068 washington reuters fbi director james comey c...
22498 anyone following the current election has ever...
14658 los angeles reuters as the us presidential el...
15236 washington reuters a single short electronic ...
Name: text, Length: 8980, dtype: object>
```

```
In [71]: y_train.head()
```

```
Out[71]: 36335 fake
12384 true
24419 fake
24740 fake
27039 true
Name: target, dtype: object
```

Decision Tree Model Classifier

```
In [72]: pipeline = Pipeline([('vect', CountVectorizer()),
                              ('tfidf', TfidfTransformer()),
                              ('model', DecisionTreeClassifier(criterion= 'entropy',
                                                                max_depth = 20,
                                                                splitter='best',
                                                                random_state=42))])

decision_tree_model = pipeline.fit(X_train, y_train)

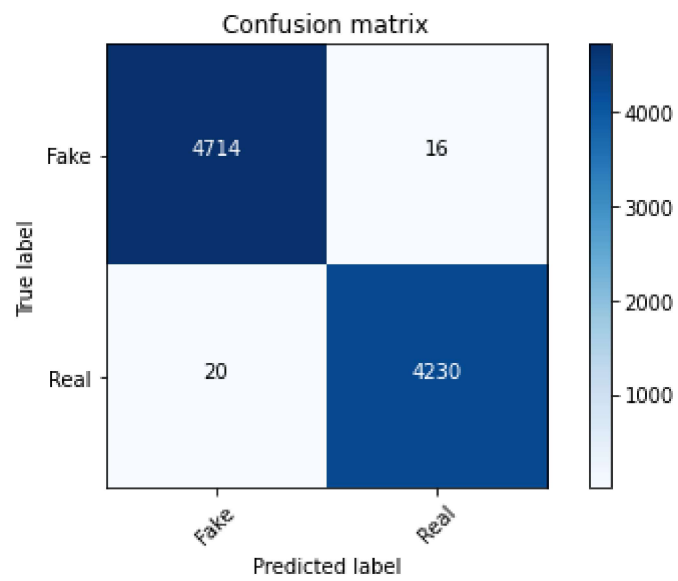
# filename = 'finalized_model.sav'
# pickle.dump(model, open(filename, 'wb'))

model_prediction = decision_tree_model.predict(X_test)
print("Decision Tree Model Accuracy: {}".format(round(accuracy_score(y_test, model_prediction)*100,2)))
```

Decision Tree Model Accuracy: 99.6%

```
In [73]: confusion_matrix = metrics.confusion_matrix(y_test, model_prediction)
plot_confusion_matrix_from_data(confusion_matrix, classes=['Fake', 'Real'])
```

Confusion matrix, without normalization



In []:

In []: