

# Objective

To use logistic regression and natural language processing (NLP) techniques such as TF-IDF to classify news articles as fake or real, using a labeled dataset from Kaggle.

## Import Libraries

```
In [1]: # Data handling
import pandas as pd
import numpy as np

# Text cleaning
import re
import string

# Visualization
import matplotlib.pyplot as plt
import seaborn as sns

# Machine Learning
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
```

## Load Data

```
In [2]: df_fake = pd.read_csv("C:\\Users\\LENOVO\\Downloads\\archive (6)\\Fake.csv")
df_true = pd.read_csv("C:\\Users\\LENOVO\\Downloads\\archive (6)\\True.csv")
```

## EDA

```
In [3]: df_fake.head()
```

```
Out[3]:
```

	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
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017

```
In [4]: df_true.head()
```

```
Out[4]:
```

	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
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017

```
In [5]: df_fake.info()  
df_fake.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 23481 entries, 0 to 23480  
Data columns (total 4 columns):  
#   Column   Non-Null Count  Dtype  
---  ---      -  
0   title    23481 non-null  object  
1   text     23481 non-null  object  
2   subject  23481 non-null  object  
3   date     23481 non-null  object  
dtypes: object(4)  
memory usage: 733.9+ KB  
Out[5]: title      0  
text      0  
subject   0  
date      0  
dtype: int64
```

```
In [6]: df_true.info()  
df_true.isnull().sum()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 21417 entries, 0 to 21416  
Data columns (total 4 columns):  
#   Column   Non-Null Count  Dtype  
---  ---      -  
0   title    21417 non-null  object  
1   text     21417 non-null  object  
2   subject  21417 non-null  object  
3   date     21417 non-null  object  
dtypes: object(4)  
memory usage: 669.4+ KB  
Out[6]: title      0  
text      0  
subject   0  
date      0  
dtype: int64
```

```
In [7]: df_fake["class"] = 0
df_true["class"] = 1
```

```
In [8]: df_fake.shape
df_true.shape
```

```
Out[8]: (21417, 5)
```

## Merging True and Fake Dataframes

```
In [9]: df_merge = pd.concat([df_fake, df_true], axis = 0 )
df_merge.head(10)
```

```
Out[9]:
```

	title	text	subject	date	class
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0
5	Racist Alabama Cops Brutalize Black Boy While...	The number of cases of cops brutalizing and ki...	News	December 25, 2017	0
6	Fresh Off The Golf Course, Trump Lashes Out A...	Donald Trump spent a good portion of his day a...	News	December 23, 2017	0
7	Trump Said Some INSANELY Racist Stuff Inside ...	In the wake of yet another court decision that...	News	December 23, 2017	0
8	Former CIA Director Slams Trump Over UN Bully...	Many people have raised the alarm regarding th...	News	December 22, 2017	0
9	WATCH: Brand-New Pro-Trump Ad Features So Muc...	Just when you might have thought we d get a br...	News	December 21, 2017	0

```
In [10]: df_merge.columns
```

```
Out[10]: Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')
```

```
In [11]: df = df_merge.drop(["title", "subject", "date"], axis = 1)
```

```
In [12]: df.isnull().sum()
```

```
Out[12]: text      0
class      0
dtype: int64
```

```
In [13]: df = df.sample(frac = 1)
df.head()
```

```
Out[13]:
```

	text	class
12710	LANGUAGE WARNING!	0
14589	Don t worry, the US government will rely on th...	0
16934	NAIROBI (Reuters) - Kenya s police said on Fri...	1
2428	Angry constituents have inundated town hall me...	0
3393	The Electoral College should tell Donald Trump...	0

```
In [14]: df.reset_index(inplace = True)
df.drop(["index"], axis = 1, inplace = True)
```

```
In [15]: df.columns
```

```
Out[15]: Index(['text', 'class'], dtype='object')
```

```
In [16]: df.head()
```

```
Out[16]:
```

	text	class
0	LANGUAGE WARNING!	0
1	Don t worry, the US government will rely on th...	0
2	NAIROBI (Reuters) - Kenya s police said on Fri...	1
3	Angry constituents have inundated town hall me...	0
4	The Electoral College should tell Donald Trump...	0

```
In [17]: def wordopt(text):
text = text.lower()
text = re.sub('[.*?\]', '', text)
text = re.sub("\W", " ", text)
text = re.sub('https?://\S+|\S+', '', text)
text = re.sub('<.*?>+', '', text)
text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
text = re.sub('\n', '', text)
text = re.sub('\w*\d\w*', '', text)
return text
```

```
In [18]: df["text"] = df["text"].apply(wordopt)
```

```
In [19]: x = df["text"]
y = df["class"]
```

## Splitting Training and Testing

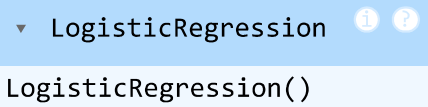
```
In [20]: x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=
```

## Convert text to vectors

```
In [21]: vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)
```

## Modeling with Logistic Regression

```
In [22]: LR = LogisticRegression()
LR.fit(xv_train, y_train)
```

```
Out[22]: 
LogisticRegression()
```

## Make Predictions

```
In [23]: pred_lr=LR.predict(xv_test)
```

## Model Evaluation

```
In [24]: LR.score(xv_test, y_test)
```

```
Out[24]: 0.9870824053452116
```

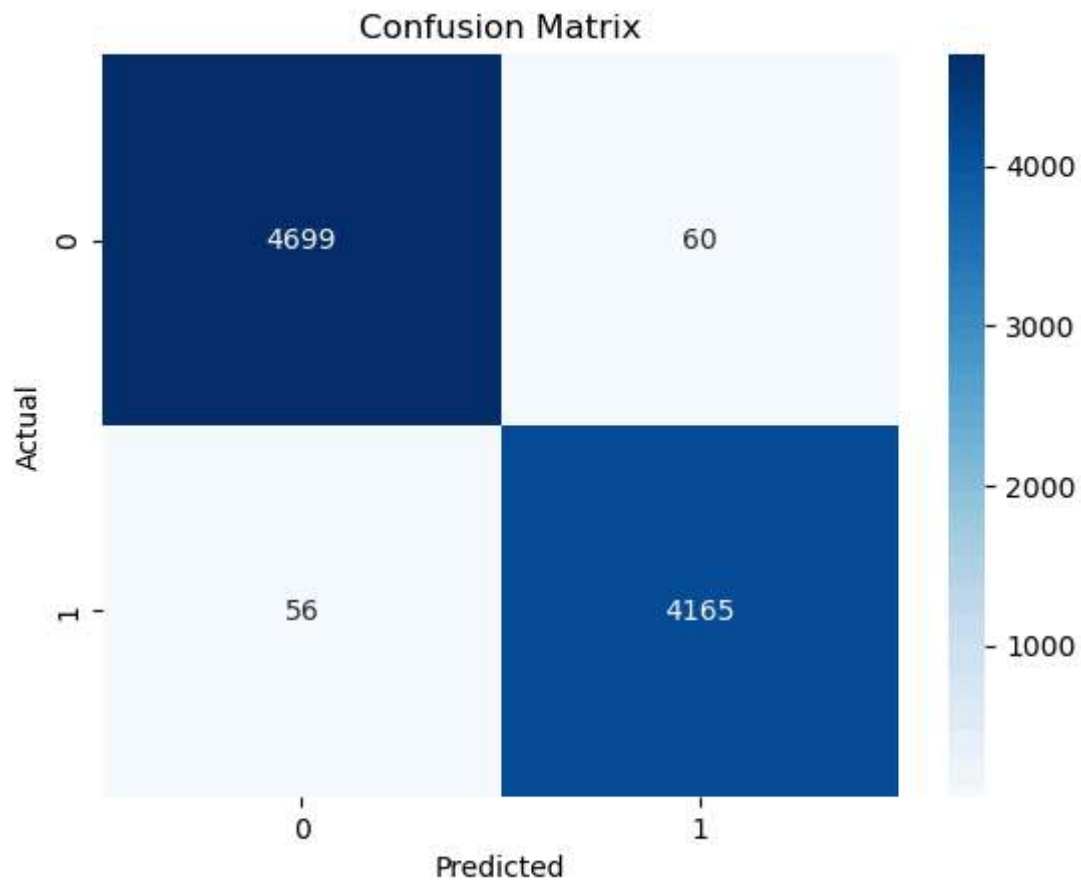
```
In [25]: print(classification_report(y_test, pred_lr))
```

	precision	recall	f1-score	support
0	0.99	0.99	0.99	4759
1	0.99	0.99	0.99	4221
accuracy			0.99	8980
macro avg	0.99	0.99	0.99	8980
weighted avg	0.99	0.99	0.99	8980

```
In [26]: print("Confusion Matrix:\n", confusion_matrix(y_test, pred_lr))
```

```
Confusion Matrix:
[[4699  60]
 [  56 4165]]
```

```
In [27]: sns.heatmap(confusion_matrix(y_test, pred_lr), annot=True, fmt='d', cmap='Blues')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```



## Conclusion

- Logistic regression achieved 98.7% accuracy on test data.
- The model can effectively detect fake vs real news using text content.
- TF-IDF helps extract meaningful features from text.