Project Title: Fake News Detection

Objective
To build a machine learning model that can accurately classify whether a given news article is **fake** or **real**, based on textual data.

Datasets Used
- **Fake.csv** – Contains fabricated news articles.
- **True.csv** – Contains authentic news articles.

Data Preprocessing
- Combined both datasets and labeled them:
    - 1 for fake news
    - 0 for real news
- Removed:
    - Punctuation
    - Stop words
    - Special characters using re and string libraries
- Applied **text cleaning** to normalize the data.

Model Building
- **Machine Learning Model Used:** Logistic Regression (possibly others too—I'll confirm after parsing further)
- Libraries:
    - sklearn for model training and evaluation
- **Data Split:**
    - train_test_split() used to divide data into training and test sets

Evaluation Metrics
- Accuracy Score
- Classification Report:
    - Precision
    - Recall
    - F1-score

Results
The notebook computes and displays:
- Model performance on test data
- Confusion matrix (if included)
- Classification metrics to show how well the model performs in detecting fake news

Conclusion
The implemented model is capable of detecting fake news based on textual data. Logistic Regression performs well after preprocessing, but further improvements (e.g., using advanced NLP models like TF-IDF + SVM or BERT) could enhance accuracy.

```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import classification_report
import re
import string


df_fake = pd.read_csv("/content/Fake.csv")
df_true = pd.read_csv("/content/True.csv")


df_fake.head(5)
```

	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

Next steps: 〔 Generate code with `df_fake` 〕 〔 ☐ View recommended plots 〕 〔 New interactive sheet 〕

```python
df_true.head(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
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

Next steps: (Generate code with `df_true`)   (☐ View recommended plots)   (New interactive sheet)

```python
df_fake["class"] = 0
df_true["class"] = 1


df_fake.shape, df_true.shape
```

→ ((23481, 5), (21417, 5))

```python
df_fake_manual_testing = df_fake.tail(10)
for i in range(23480,23470,-1):
    df_fake.drop([i], axis = 0, inplace = True)


df_true_manual_testing = df_true.tail(10)
for i in range(21416,21406,-1):
    df_true.drop([i], axis = 0, inplace = True)


df_fake.shape, df_true.shape
```

→ ((23471, 5), (21407, 5))

```python
df_fake_manual_testing["class"] = 0
df_true_manual_testing["class"] = 1
```

→ /tmp/ipython-input-10-860779283.py:1: SettingWithCopyWarning:
    A value is trying to be set on a copy of a slice from a DataFrame.
    Try using .loc[row_indexer,col_indexer] = value instead

    See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
      df_fake_manual_testing["class"] = 0

```
/tmp/ipython-input-10-860779283.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
  df_true_manual_testing["class"] = 1
```

df_fake_manual_testing.head(10)

	title	text	subject	date	class
23471	Seven Iranians freed in the prisoner swap have...	21st Century Wire says This week, the historic...	Middle-east	January 20, 2016	0
23472	#Hashtag Hell & The Fake Left	By Dady Chery and Gilbert MercierAll writers ...	Middle-east	January 19, 2016	0
23473	Astroturfing: Journalist Reveals Brainwashing ...	Vic Bishop Waking TimesOur reality is carefull...	Middle-east	January 19, 2016	0
23474	The New American Century: An Era of Fraud	Paul Craig RobertsIn the last years of the 20t...	Middle-east	January 19, 2016	0
23475	Hillary Clinton: 'Israel First' (and no peace ...	Robert Fantina CounterpunchAlthough the United...	Middle-east	January 18, 2016	0
23476	McPain: John McCain Furious That Iran Treated ...	21st Century Wire says As 21WIRE reported earl...	Middle-east	January 16, 2016	0
23477	JUSTICE? Yahoo Settles E-mail Privacy Class-ac...	21st Century Wire says It s a familiar theme. ...	Middle-east	January 16, 2016	0
23478	Sunnistan: US and Allied 'Safe Zone' Plan to T...	Patrick Henningsen 21st Century WireRemember ...	Middle-east	January 15, 2016	0
23479	How to Blow $700 Million: Al Jazeera America F...	21st Century Wire says Al Jazeera America will...	Middle-east	January 14, 2016	0
23480	10 U.S. Navy Sailors Held by Iranian Military ...	21st Century Wire says As 21WIRE predicted in ...	Middle-east	January 12, 2016	0

df_true_manual_testing.head(10)

	title	text	subject	date	class
21407	Mata Pires, owner of embattled Brazil builder ...	SAO PAULO (Reuters) - Cesar Mata Pires, the ow...	worldnews	August 22, 2017	1
21408	U.S., North Korea clash at U.N. forum over nuc...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21409	U.S., North Korea clash at U.N. arms forum on ...	GENEVA (Reuters) - North Korea and the United ...	worldnews	August 22, 2017	1
21410	Headless torso could belong to submarine journ...	COPENHAGEN (Reuters) - Danish police said on T...	worldnews	August 22, 2017	1
21411	North Korea shipments to Syria chemical arms a...	UNITED NATIONS (Reuters) - Two North Korean sh...	worldnews	August 21, 2017	1
21412	'Fully committed' NATO backs new U.S. approach...	BRUSSELS (Reuters) - NATO allies on Tuesday we...	worldnews	August 22, 2017	1
21413	LexisNexis withdrew two products from Chinese ...	LONDON (Reuters) - LexisNexis, a provider of I...	worldnews	August 22, 2017	1
21414	Minsk cultural hub becomes haven from authorities	MINSK (Reuters) - In the shadow of disused Sov...	worldnews	August 22, 2017	1
21415	Vatican upbeat on possibility of Pope Francis ...	MOSCOW (Reuters) - Vatican Secretary of State ...	worldnews	August 22, 2017	1
21416	Indonesia to buy $1.14 billion worth of Russia...	JAKARTA (Reuters) - Indonesia will buy 11 Sukh...	worldnews	August 22, 2017	1

```python
df_manual_testing = pd.concat([df_fake_manual_testing,df_true_manual_testing], axis = 0)
df_manual_testing.to_csv("manual_testing.csv")
```

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

	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

Next steps: Generate code with `df_merge`   View recommended plots   New interactive sheet

```
df_merge.columns
```

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

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

```
df.isnull().sum()
```

	0
text	0
class	0

dtype: int64

```
df = df.sample(frac = 1)
```

(▌ Code) (▌ Text)

```
df.head()
```

	text	class
5487	WASHINGTON (Reuters) - Despite his campaign vo...	1
19858	MANILA (Reuters) - Philippine President Rodrig...	1
19670	Comedian Dave Chappelle has been prepping for ...	0
1661	Donald Trump is quickly closing in on his firs...	0
2801	Donald Trump s latest executive orders are bei...	0

Next steps: (Generate code with df) (▢ View recommended plots) (New interactive sheet)

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

```
df.columns
```

    Index(['text', 'class'], dtype='object')

```
df.head()
```

	text	class
0	WASHINGTON (Reuters) - Despite his campaign vo...	1
1	MANILA (Reuters) - Philippine President Rodrig...	1
2	Comedian Dave Chappelle has been prepping for ...	0
3	Donald Trump is quickly closing in on his firs...	0
4	Donald Trump s latest executive orders are bei...	0

Next steps: (Generate code with df) (▢ View recommended plots) (New interactive sheet)

```
def wordopt(text):
    text = text.lower()
    text = re.sub('\[.*?\]', '', text)
```

```python
    text = re.sub("\\W"," ",text)
    text = re.sub('https?://\S+|www\.\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
```

```python
df["text"] = df["text"].apply(wordopt)
```

```python
x = df["text"]
y = df["class"]
```

```python
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25)
```

```python
from sklearn.feature_extraction.text import TfidfVectorizer

vectorization = TfidfVectorizer()
xv_train = vectorization.fit_transform(x_train)
xv_test = vectorization.transform(x_test)
```

```python
from sklearn.linear_model import LogisticRegression

LR = LogisticRegression()
LR.fit(xv_train,y_train)
```

```
▼ LogisticRegression  ⓘ ?

LogisticRegression()
```

```python
pred_lr=LR.predict(xv_test)
```

```python
LR.score(xv_test, y_test)
```

```
0.9866310160427807
```

```python
print(classification_report(y_test, pred_lr))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      5828
           1       0.99      0.99      0.99      5392

    accuracy                           0.99     11220
   macro avg       0.99      0.99      0.99     11220
weighted avg       0.99      0.99      0.99     11220
```

```python
from sklearn.tree import DecisionTreeClassifier

DT = DecisionTreeClassifier()
DT.fit(xv_train, y_train)
```

```
       ▾  DecisionTreeClassifier   ⓘ �ⓘ
     DecisionTreeClassifier()
```

```python
pred_dt = DT.predict(xv_test)
```

```python
DT.score(xv_test, y_test)
```

```
0.9966131907308378
```

```python
print(classification_report(y_test, pred_dt))
```

```
                 precision    recall  f1-score   support

             0       1.00      1.00      1.00      5828
             1       1.00      1.00      1.00      5392

      accuracy                           1.00     11220
     macro avg       1.00      1.00      1.00     11220
  weighted avg       1.00      1.00      1.00     11220
```

```python
from sklearn.ensemble import GradientBoostingClassifier

GBC = GradientBoostingClassifier(random_state=0)
GBC.fit(xv_train, y_train)
```

```
       ▾     GradientBoostingClassifier   ⓘ ⓘ
     GradientBoostingClassifier(random_state=0)
```

```python
pred_gbc = GBC.predict(xv_test)
```

```python
GBC.score(xv_test, y_test)
```

```
0.9959893048128342
```

```python
print(classification_report(y_test, pred_gbc))
```

```
                 precision    recall  f1-score   support

             0       1.00      0.99      1.00      5828
             1       0.99      1.00      1.00      5392

      accuracy                           1.00     11220
     macro avg       1.00      1.00      1.00     11220
```

```
    weighted avg          1.00        1.00        1.00        11220
```

```python
from sklearn.ensemble import RandomForestClassifier

RFC = RandomForestClassifier(random_state=0)
RFC.fit(xv_train, y_train)
```

```
        ▼       RandomForestClassifier        ⓘ ⓘ
        RandomForestClassifier(random_state=0)
```

```python
pred_rfc = RFC.predict(xv_test)
```

```python
RFC.score(xv_test, y_test)
```

```
0.9899286987522281
```

```python
print(classification_report(y_test, pred_rfc))
```

```
                  precision    recall  f1-score   support

             0        0.99      0.99      0.99      5828
             1        0.99      0.99      0.99      5392

      accuracy                            0.99     11220
     macro avg        0.99      0.99      0.99     11220
  weighted avg        0.99      0.99      0.99     11220
```

```python
def output_lable(n):
    if n == 0:
        return "Fake News"
    elif n == 1:
        return "Not A Fake News"

def manual_testing(news):
    testing_news = {"text":[news]}
    new_def_test = pd.DataFrame(testing_news)
    new_def_test["text"] = new_def_test["text"].apply(wordopt)
    new_x_test = new_def_test["text"]
    new_xv_test = vectorization.transform(new_x_test)
    pred_LR = LR.predict(new_xv_test)
    pred_DT = DT.predict(new_xv_test)
    pred_GBC = GBC.predict(new_xv_test)
    pred_RFC = RFC.predict(new_xv_test)

    return print("\n\nLR Prediction: {} \nDT Prediction: {} \nGBC Prediction: {} \nRFC Pr
```

```
news = str(input())
manual_testing(news)
```

BRUSSELS (Reuters) - NATO allies on Tuesday welcomed President Donald Trump s decisio

```
    LR Prediction: Not A Fake News
    DT Prediction: Not A Fake News
    GBC Prediction: Not A Fake News
    RFC Prediction: Not A Fake News
```

```
news = str(input())
manual_testing(news)
```

Vic Bishop Waking TimesOur reality is carefully constructed by powerful corporate, po

```
    LR Prediction: Fake News
    DT Prediction: Fake News
    GBC Prediction: Fake News
    RFC Prediction: Fake News
```

```
news = str(input())
manual_testing(news)
```

SAO PAULO (Reuters) - Cesar Mata Pires, the owner and co-founder of Brazilian enginee