Fake News Detection using Machine Learning

Importing libraries

```
In [1]:  import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt
         %matplotlib inline
         import re
         import string
         from sklearn.model_selection import train_test_split
         from sklearn.metrics import accuracy_score
         from sklearn.metrics import confusion_matrix
         from sklearn.metrics import classification_report
```

Load Dataset

```
In [2]:  # fake and Real Dataset
         data_fake=pd.read_csv(r"C:\Users\hp\OneDrive\Documents\Desktop\Fake.csv")
         data_true=pd.read_csv(r"C:\Users\hp\OneDrive\Documents\Desktop\True.csv")
```

```
In [3]:  data_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]: ▶| `data_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

Creating new column called "class" for Fake News and Real News dataset to categories fake and true news

In [5]: ▶|
```
data_fake["class"] = 0
data_true["class"] = 1
```

In [6]: ▶| `data_fake.head()`

Out[6]:

	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

In [7]: ▶| `data_fake.shape`

Out[7]: `(23481, 5)`

In [8]: ▶| `data_true.head()`

Out[8]:

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

In [9]: ▶| `data_true.shape`

Out[9]: `(21417, 5)`

Removing last 10 rows from both the dataset, for manual testing

In [10]: ▶|
```python
data_fake_manual_testing = data_fake.tail(10)
for i in range(23480, 23470, -1):
    if i in data_fake.index:
        data_fake.drop([i], axis=0, inplace=True)

data_true_manual_testing = data_true.tail(10)
for i in range(21416, 21406, -1):
    if i in data_true.index:
        data_true.drop([i], axis=0, inplace=True)
```

In [11]: ▶| `data_fake.shape, data_true.shape`

Out[11]: `((23471, 5), (21407, 5))`

In [12]: ▶| `data_fake_manual_testing.head()`

Out[12]:

	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

In [13]: ▶| `data_true_manual_testing.head(10)`

Out[13]:

	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

marging the manual testing dataframe in single dataset and save it in a csv file

In [14]: ▶|
```
data_manual_testing = pd.concat([data_fake_manual_testing,data_true_manual_testing],axis = 0
data_manual_testing.to_csv("manual_testing_dataset.csv")
```

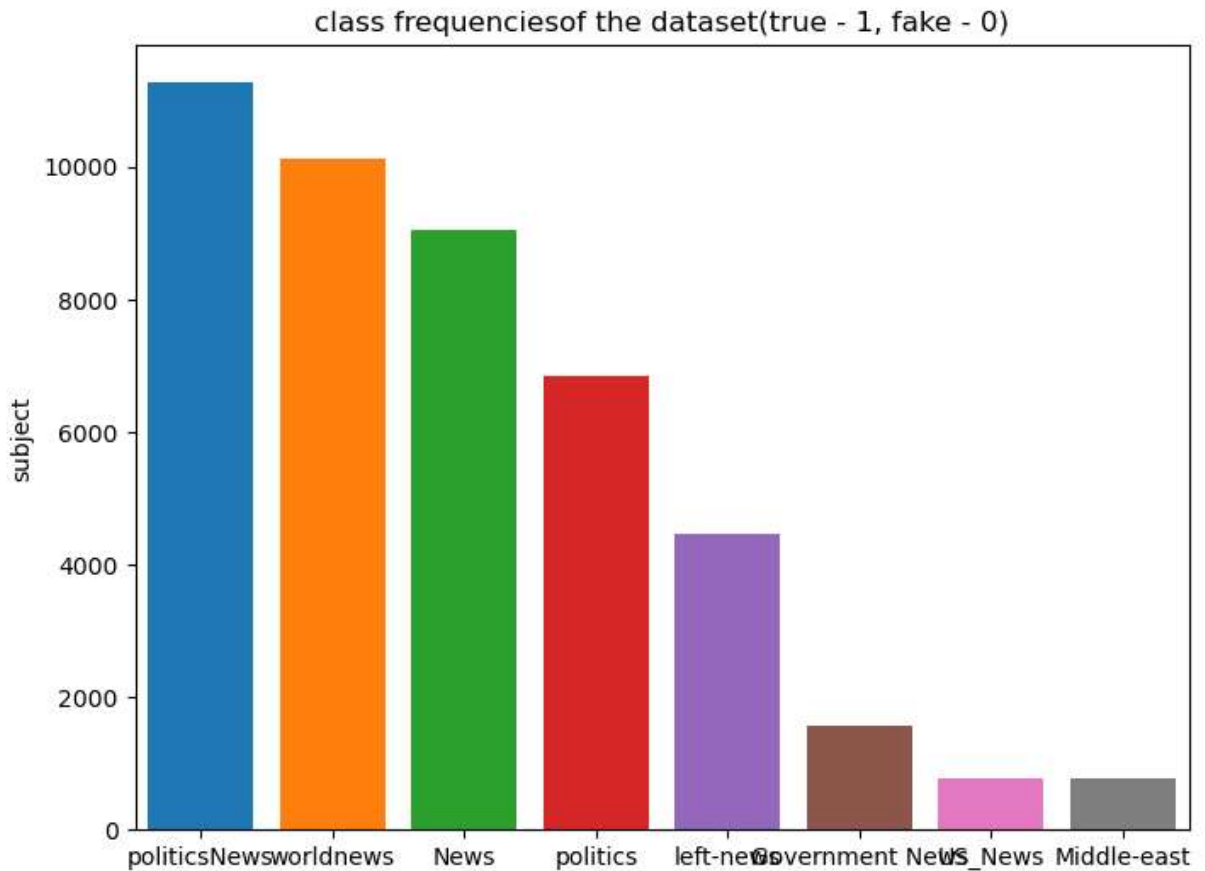# Marging the main fake and true DataFeame

In [15]: ▶|
```
data_marge = pd.concat([data_fake, data_true], axis =0)
data_marge.head()
```

Out[15]:

	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

```
In [16]:    ▶|  plt.figure(figsize = [8,6], clear = True, facecolor = "white")
                sns.barplot(x = data_marge["subject"].value_counts().index,
                            y = data_marge["subject"].value_counts(),
                            saturation = 1).set(title = "class frequenciesof the dataset(true - 1, fake - 0)"
```

class frequenciesof the dataset(true - 1, fake - 0)


```
In [17]:    ▶|  data_marge.columns
```

Out[17]:   Index(['title', 'text', 'subject', 'date', 'class'], dtype='object')

"title","subject" and "Date" columns is not required for detecting the fake news so we have to drop these columns.

```
In [18]:    ▶|  data = data_marge.drop(["title", "subject", "date"], axis = 1)
                data.head()
```

Out[18]:

	text	class
0	Donald Trump just couldn t wish all Americans ...	0
1	House Intelligence Committee Chairman Devin Nu...	0
2	On Friday, it was revealed that former Milwauk...	0
3	On Christmas day, Donald Trump announced that ...	0
4	Pope Francis used his annual Christmas Day mes...	0

```
In [19]:    ▶|  data.isnull().sum()
```

Out[19]:   text   0
            class   0
            dtype: int64

Randomly shuffling the dataframe

```
In [20]:    ▶| data = data.sample(frac = 1)
```

```
In [21]:    ▶| data.head()
```

Out[21]:

	text	class
19404	(This version of September 20 s story correct...	1
13789	When will the media report on this nice group ...	0
22680	Patrick Henningsen 21st Century Wire UPDATE: 1...	0
14433	HARARE (Reuters) - Zimbabwe President Robert M...	1
12221	Talk radio hosts went nuts with the news of Go...	0

Reseting index

```
In [22]:    ▶| data.reset_index(inplace = True)
               data.drop(["index"], axis = 1, inplace = True)
```

```
In [23]:    ▶| data.columns
```

Out[23]: Index(['text', 'class'], dtype='object')

```
In [24]:    ▶| data.head(10)
```

Out[24]:

	text	class
0	(This version of September 20 s story correct...	1
1	When will the media report on this nice group ...	0
2	Patrick Henningsen 21st Century Wire UPDATE: 1...	0
3	HARARE (Reuters) - Zimbabwe President Robert M...	1
4	Talk radio hosts went nuts with the news of Go...	0
5	TWO PROTAGONISTS: Jesus Campos, and alleged sh...	0
6	GENERALLY, VENEZUELANS DO NOT EAT RABBITS AND...	0
7	WASHINGTON (Reuters) - President Donald Trump ...	1
8	WASHINGTON (Reuters) - Secretary of State John...	1
9	MEXICO CITY (Reuters) - At first, Mexico's gov...	1

creating a function to convert the text in lowercase, remove the extra space , special cheracter, url and links

In [25]: ▶
```python
def word_drop(text):
    text = text.lower()
    text = re.sub('\[.*?\]', '', text)
    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
```

In [26]: ▶
```python
data["text"] = data["text"].apply(word_drop)
```

In [27]: ▶
```python
data.head()
```

Out[27]:

	text	class
0	this version of september s story corrects ...	1
1	when will the media report on this nice group ...	0
2	patrick henningsen century wire update la...	0
3	harare reuters zimbabwe president robert m...	1
4	talk radio hosts went nuts with the news of go...	0

Creating Defining dependent and independent variable as x and y

In [28]: ▶
```python
x = data["text"]
y = data["class"]
```

Splitting the dataset into training set and testing set.

In [29]: ▶
```python
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.30, random_state = 1
```

Converting "text" into vectors

In [30]: ▶
```python
from sklearn.feature_extraction.text import TfidfVectorizer
```

In [31]: ▶
```python
vectorization = TfidfVectorizer()
x_vec_train = vectorization.fit_transform(x_train)
x_vec_test = vectorization.transform(x_test)
```

Logistic Regression

In [32]: ▶
```python
from sklearn.linear_model import LogisticRegression
log = LogisticRegression()
```

In [33]: ▶| `log.fit(x_vec_train,y_train)`

Out[33]:
```
▾ LogisticRegression
LogisticRegression()
```

In [34]: ▶|
```
pred_log = log.predict(x_vec_test)
pred_log
```

Out[34]: `array([0, 0, 1, ..., 1, 0, 0], dtype=int64)`

In [35]: ▶| `log.score(x_vec_test,y_test)`
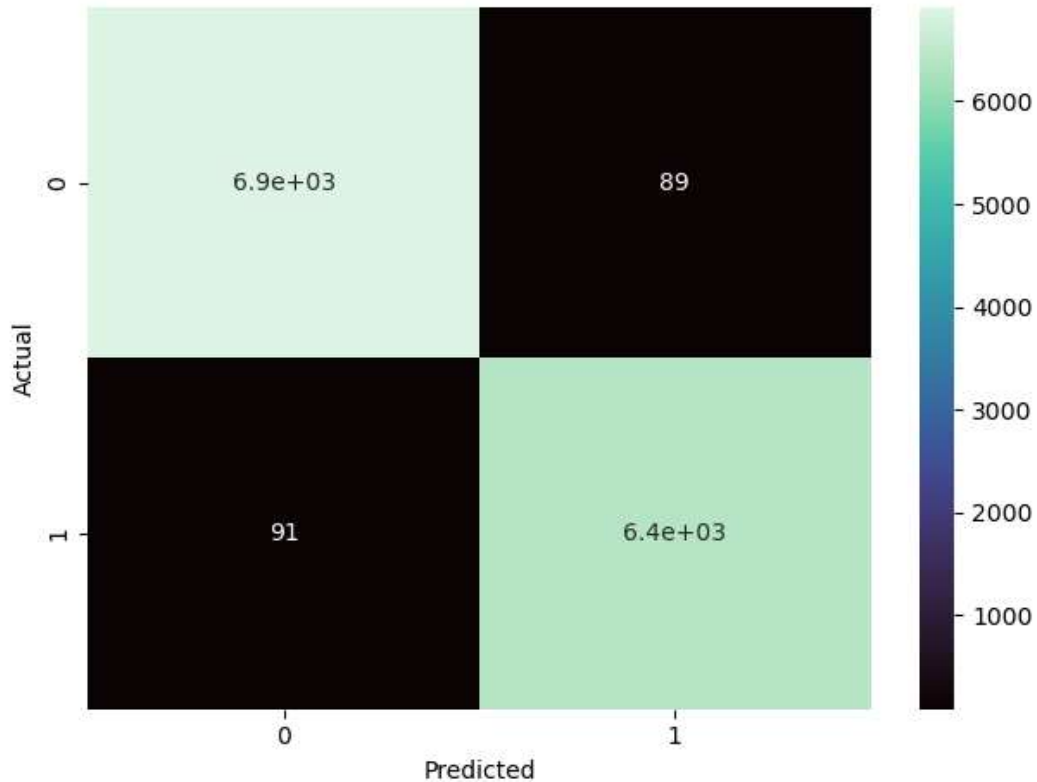
Out[35]: `0.9866310160427807`

In [36]: ▶|
```
con_log = confusion_matrix(y_test,pred_log)
con_log
```

Out[36]:
```
array([[6913,   89],
       [  91, 6371]], dtype=int64)
```

In [37]: ▶| `print(classification_report(y_test,pred_log))`

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      7002
           1       0.99      0.99      0.99      6462

    accuracy                           0.99     13464
   macro avg       0.99      0.99      0.99     13464
weighted avg       0.99      0.99      0.99     13464
```

```
In [38]:    ▶|  sns.heatmap(pd.DataFrame(con_log), annot=True, cmap="mako")
                plt.xlabel("Predicted")
                plt.ylabel("Actual")
                plt.tight_layout()
```


Decision Tree Classification

```
In [39]:    ▶|  from sklearn.tree import DecisionTreeClassifier
```

```
In [40]:    ▶|  dt = DecisionTreeClassifier()
                dt.fit(x_vec_train, y_train)
```

Out[40]:
```
▼ DecisionTreeClassifier
  DecisionTreeClassifier()
```

```
In [41]:    ▶|  pred_dt = dt.predict(x_vec_test)
                pred_dt
```

Out[41]: array([0, 0, 1, ..., 1, 0, 0], dtype=int64)

```
In [42]:    ▶|  dt.score(x_vec_test, y_test)
```

Out[42]: 0.9955436720142602

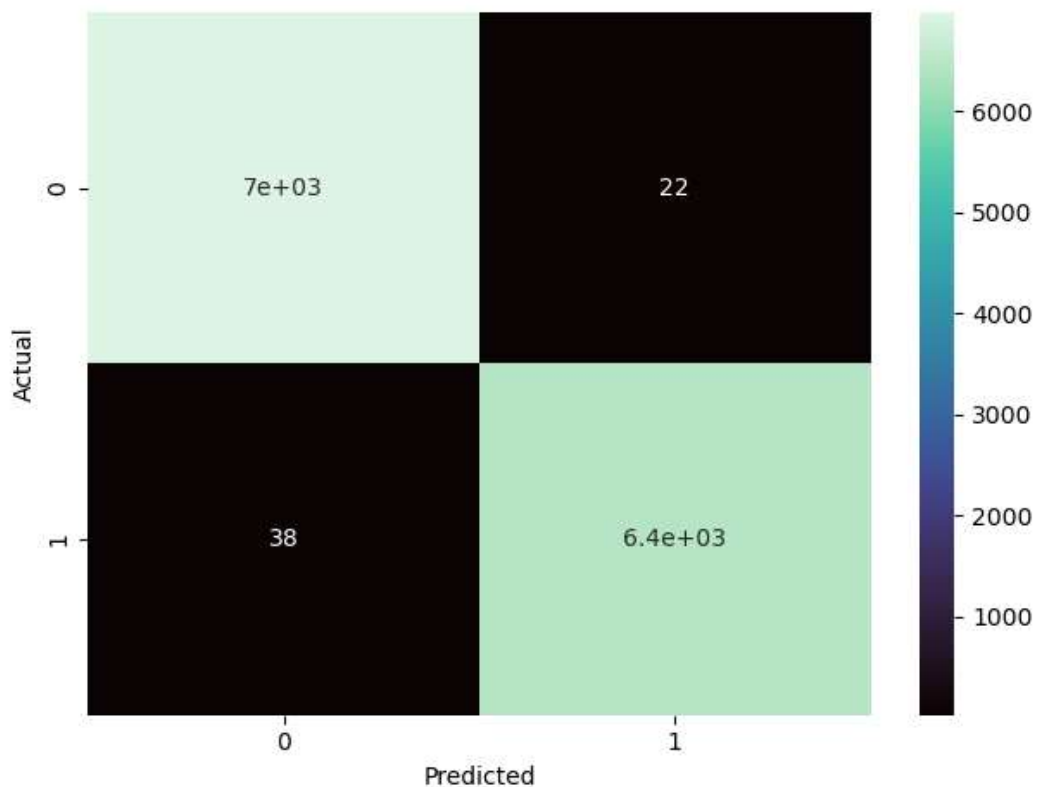In [43]: ▶|
```python
con_dt = confusion_matrix(y_test,pred_dt)
con_dt
```

Out[43]:
```
array([[6980,   22],
       [  38, 6424]], dtype=int64)
```

In [44]: ▶|
```python
print(classification_report(y_test, pred_dt))
```

```
              precision    recall  f1-score   support

           0       0.99      1.00      1.00      7002
           1       1.00      0.99      1.00      6462

    accuracy                           1.00     13464
   macro avg       1.00      1.00      1.00     13464
weighted avg       1.00      1.00      1.00     13464
```

In [45]: ▶|
```python
sns.heatmap(pd.DataFrame(con_dt), annot=True, cmap="mako")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.tight_layout()
```


Random Forest Classifier

In [46]: ▶|
```python
from sklearn.ensemble import RandomForestClassifier
```

In [47]: ► ```
rfc = RandomForestClassifier(random_state=0)
rfc.fit(x_vec_train, y_train)
```

Out[47]:
```
▼   RandomForestClassifier
RandomForestClassifier(random_state=0)
```

In [48]: ► ```
pred_rfc = rfc.predict(x_vec_test)
```

In [49]: ► ```
rfc.score(x_vec_test, y_test)
```

Out[49]: 0.9894533571004159

In [50]: ► ```
con_rfc = confusion_matrix(y_test,pred_rfc)
con_rfc
```
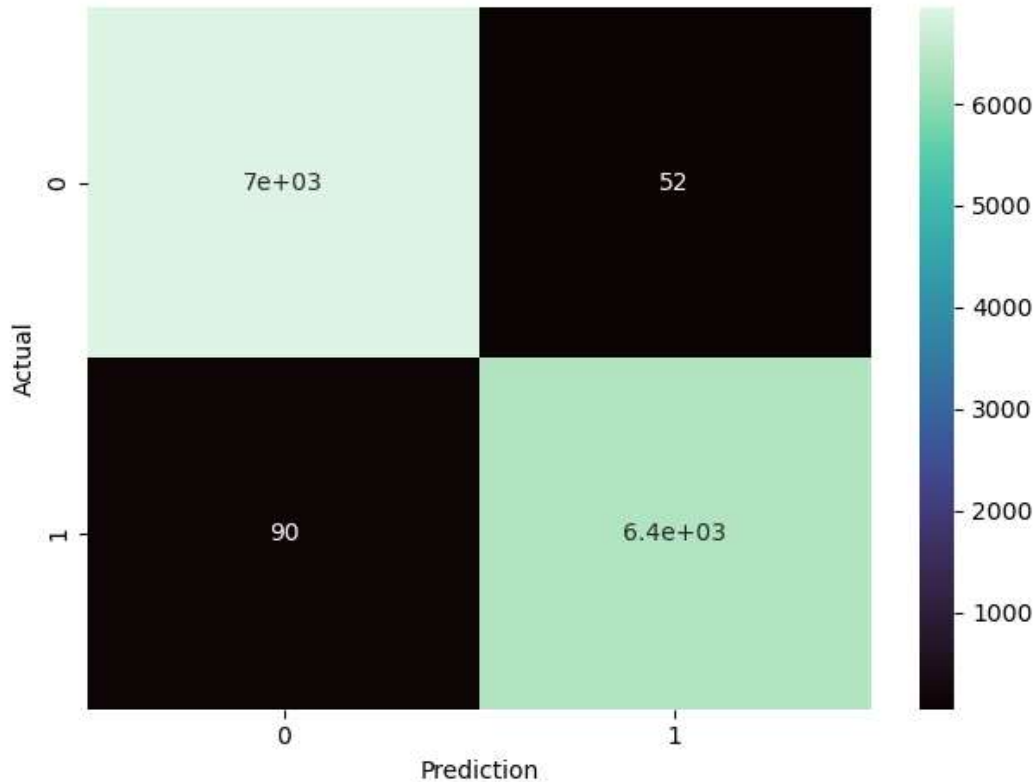
Out[50]:
```
array([[6950,   52],
       [  90, 6372]], dtype=int64)
```

In [51]: ► ```
print(classification_report(y_test, pred_rfc))
```

```
              precision    recall  f1-score   support

           0       0.99      0.99      0.99      7002
           1       0.99      0.99      0.99      6462

    accuracy                           0.99     13464
   macro avg       0.99      0.99      0.99     13464
weighted avg       0.99      0.99      0.99     13464
```

In [52]: ▶|
```python
sns.heatmap(pd.DataFrame(con_rfc), annot=True, cmap="mako")
plt.xlabel("Prediction")
plt.ylabel("Actual")
plt.tight_layout()
```


Gradient Boosting Classifier

In [53]: ▶|
```python
from sklearn.ensemble import GradientBoostingClassifier
```

In [54]: ▶|
```python
gbc = GradientBoostingClassifier(random_state=0)
gbc.fit(x_vec_train, y_train)
```

Out[54]:
```
        ▾          GradientBoostingClassifier
GradientBoostingClassifier(random_state=0)
```

In [55]: ▶|
```python
pred_gbc = gbc.predict(x_vec_test)
```

In [56]: ▶|
```python
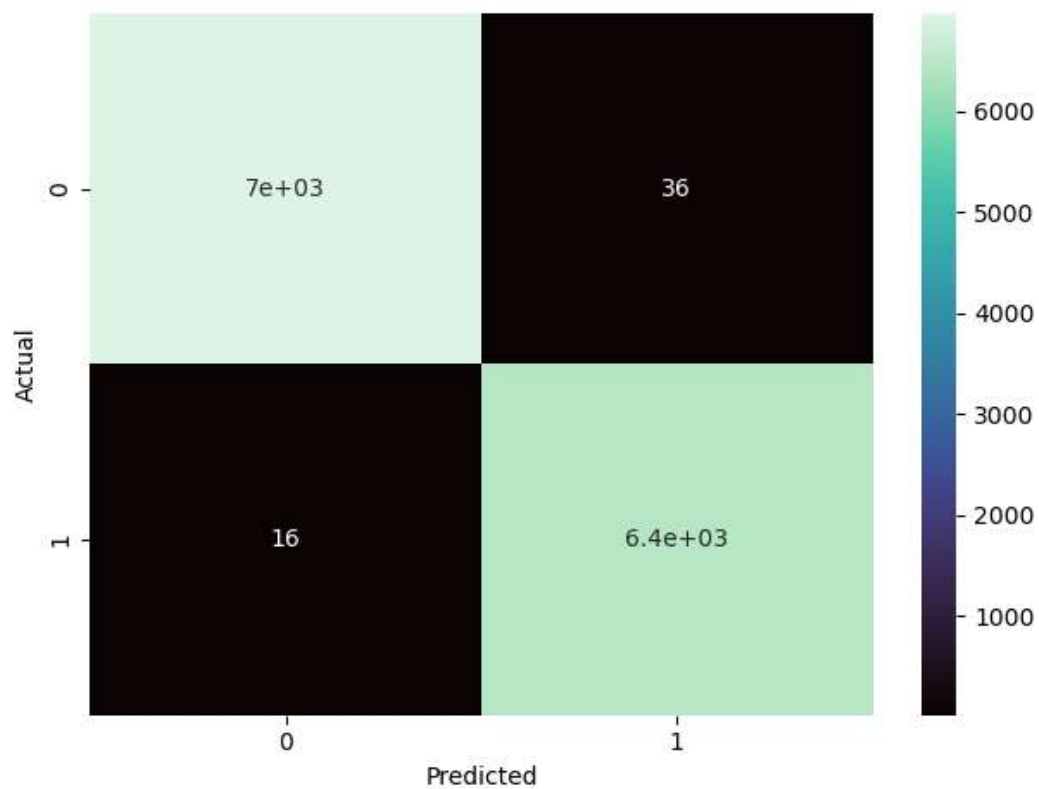gbc.score(x_vec_test, y_test)
```

Out[56]: 0.9961378490790256

In [57]: ▶|
```python
con_gbc = confusion_matrix(y_test,pred_gbc)
```

In [58]: ▶| `print(classification_report(y_test, pred_gbc))`

```
              precision    recall  f1-score   support

           0       1.00      0.99      1.00      7002
           1       0.99      1.00      1.00      6462

    accuracy                           1.00     13464
   macro avg       1.00      1.00      1.00     13464
weighted avg       1.00      1.00      1.00     13464
```

In [59]: ▶|
```python
sns.heatmap(pd.DataFrame(con_gbc), annot=True, cmap="mako")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.tight_layout()
```


Model Testing With Manual Entry

In [60]: ▶|
```python
def output_lable(n):
    if n == 0:
        return "Fake News"
    elif n == 1:
        return "Real 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(word_drop)
    new_x_test = new_def_test["text"]
    new_x_vec_test = vectorization.transform(new_x_test)
    pred_LOG = log.predict(new_x_vec_test)
    pred_DT = dt.predict(new_x_vec_test)
    pred_GBC = gbc.predict(new_x_vec_test)
    pred_RFC = rfc.predict(new_x_vec_test)
    return print("\n\nLogistic Regression Prediction: {} \nDecision Tree Classification Pred
```

In [68]: ▶|
```python
news = str(input())
manual_testing(news)
```

1

Logistic Regression Prediction: Fake News
Decision Tree Classification Prediction: Fake News
Gradient Boosting Classifier Prediction: Fake News
Random Forest Classifier Prediction: Fake News

In [ ]: ▶|