```
!pip install transformers

# Importing the necessary libraries
import pandas as pd
import nltk
import re
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import WordNetLemmatizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from transformers import BertTokenizer, BertForSequenceClassification, AdamW
import torch

# Download necessary NLTK resources
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')

# Loading fake dataset
fake_data = pd.read_csv('/content/Fake.csv')
print(fake_data.info())
print(fake_data.head())

# Loading true dataset
true_data = pd.read_csv('/content/True.csv')
print(true_data.info())
print(true_data.head())

# Combine the datasets into one
fake_data['label'] = 1  # Assigning label 1 to fake news
true_data['label'] = 0  # Assigning label 0 to true news
data = pd.concat([fake_data, true_data], ignore_index=True)
print(data.info())

# Data preprocess

# Data Cleaning with regular expressions
data['text'] = data['text'].apply(lambda x: re.sub('<[^>]+>', '', x))  # Remove HTML tags
data['text'] = data['text'].apply(lambda x: re.sub('[^a-zA-Z\s]', '', x))  # Remove non-alphabetical characters

# Convert text to lowercase
data['text'] = data['text'].str.lower()

# Tokenization
data['tokens'] = data['text'].apply(word_tokenize)

# Stopword Removal
stop_words = set(stopwords.words('english'))
data['filtered_tokens'] = data['tokens'].apply(lambda tokens: [word for word in tokens if word not in stop_words])

# Text Lemmatization
lemmatizer = WordNetLemmatizer()
data['lemmatized_tokens'] = data['filtered_tokens'].apply(lambda tokens: [lemmatizer.lemmatize(word) for word in tokens])

# Text Vectorization (using TF-IDF)
tfidf_vectorizer = TfidfVectorizer(max_features=1000)  # Adjust max_features as needed
X_tfidf = tfidf_vectorizer.fit_transform(data['lemmatized_tokens'].apply(' '.join))

# Model Training and Evaluation

# Split the data
X_train, X_test, y_train, y_test = train_test_split(X_tfidf, data['label'], test_size=0.2, random_state=42)

# Naive Bayes
nb_model = MultinomialNB()
nb_model.fit(X_train, y_train)
nb_pred = nb_model.predict(X_test)
nb_accuracy = accuracy_score(y_test, nb_pred)


# Random Forest
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
rf_pred = rf_model.predict(X_test)
rf_accuracy = accuracy_score(y_test, rf_pred)

print(f'Naive Bayes Accuracy: {nb_accuracy}')
print(f'Random Forest Accuracy: {rf_accuracy}')
```

```
2  On Friday, it was revealed that former Milwauk...    News
3  On Christmas day, Donald Trump announced that ...    News
4  Pope Francis used his annual Christmas Day mes...    News

                date
0  December 31, 2017
1  December 31, 2017
2  December 30, 2017
3  December 29, 2017
4  December 25, 2017
<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
None
                                               title  \
0  As U.S. budget fight looms, Republicans flip t...   
1  U.S. military to accept transgender recruits o...   
2  Senior U.S. Republican senator: 'Let Mr. Muell...   
3  FBI Russia probe helped by Australian diplomat...   
4  Trump wants Postal Service to charge 'much mor...   

                                                text       subject  \
0  WASHINGTON (Reuters) - The head of a conservat...  politicsNews   
1  WASHINGTON (Reuters) - Transgender people will...  politicsNews   
2  WASHINGTON (Reuters) - The special counsel inv...  politicsNews   
3  WASHINGTON (Reuters) - Trump campaign adviser ...  politicsNews   
4  SEATTLE/WASHINGTON (Reuters) - President Donal...  politicsNews   

                 date  
0  December 31, 2017   
1  December 29, 2017   
2  December 31, 2017   
3  December 30, 2017   
4  December 29, 2017   
<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   label    44898 non-null  int64 
dtypes: int64(1), object(4)
memory usage: 1.7+ MB
None
Naive Bayes Accuracy: 0.9200445434298441
Random Forest Accuracy: 0.9978841870824053
```