

```
import pandas as pd
import nltk
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.naive_bayes import MultinomialNB
from sklearn.ensemble import RandomForestClassifier

# Download NLTK stopwords (if not already downloaded)
nltk.download("stopwords")

# Load the datasets into Pandas DataFrames
fake_df = pd.read_csv("Fake.csv")
true_df = pd.read_csv("True.csv")

# Add labels to the data (fake news labeled as 1, real news labeled as 0)
fake_df['label'] = 1
true_df['label'] = 0

# Combine both datasets into one
data = pd.concat([fake_df, true_df], ignore_index=True)

# Shuffle the data
data = data.sample(frac=1, random_state=42).reset_index(drop=True)

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!

# Initialize stopwords and TF-IDF Vectorizer
stop_words = set(stopwords.words("english"))
tfidf_vectorizer = TfidfVectorizer(max_features=5000)

# Preprocessing function
def preprocess_text(text):
    text = text.lower()
    text = ''.join([char if char.isalnum() or char.isspace() else ' ' for char in text])
    words = text.split()
    words = [word for word in words if word not in stop_words]
    return ' '.join(words)

# Apply preprocessing to the text column
data['text'] = data['text'].apply(preprocess_text)

# Vectorize the text using TF-IDF
X = tfidf_vectorizer.fit_transform(data['text'])

# Split the data into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, data['label'], test_size=0.2, random_state=42)

data_fake = pd.read_csv('Fake.csv')
data_true = pd.read_csv('True.csv')
```

```
data_fake.head()
```

	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		

```
data_true.head()
```

```

                                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
4      U.S. military to accept transgender recruits a...  WASHINGTON (Reuters) - Transgender recruits will...  politicsNews  December 30, 2017
data_fake["class"] = 0
data_true["class"]= 1
3      FBI Russia probe helped by Australian diplomat...  WASHINGTON (Reuters) - Trump campaign adviser ...  politicsNews  December 30, 2017
data_fake.shape, data_true.shape

((23481, 5), (21417, 5))
```

```

data_fake_manual_testing = data_fake.tail(10)
for i in range(23480,23470,-1):
    data_fake.drop([i], axis = 0, inplace = True)

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

```

data_fake.shape, data_true.shape

((23471, 5), (21407, 5))
```

```

data_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

```

data_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

```

data_merge = pd.concat([data_fake,data_true], axis = 0)
data_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	

```
data_merge.columns
```

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

```
data = data_merge.drop(['title', 'subject', 'date'], axis = 1)
```

```
data.isnull().sum()
```

```
text      0
class     0
dtype: int64
```

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

```
data.head()
```

	text	class	
9810	John McCain really doesn t care what topic rep...	0	
3706	(Reuters) - The head of the U.S. House Armed S...	1	
8706	Filmmaker Michael Moore, who was born in Flint...	0	
17009	KUALA LUMPUR (Reuters) - A senior Malaysian op...	1	

```
data.reset_index(inplace = True)
data.drop(['index'], axis = 1, inplace = True)
```

```
data.columns
```

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

```
data.head()
```

	text	class	
0	John McCain really doesn t care what topic rep...	0	
1	(Reuters) - The head of the U.S. House Armed S...	1	
2	Filmmaker Michael Moore, who was born in Flint...	0	
3	KUALA LUMPUR (Reuters) - A senior Malaysian op...	1	

```
def wordopt(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
```

```
# Initialize classifiers
classifiers = [
    ("Support Vector Machine", SVC(kernel='linear')),
    ("Logistic Regression", LogisticRegression(max_iter=1000)),
    ("Multinomial Naive Bayes", MultinomialNB()),
    ("Random Forest", RandomForestClassifier()),
]
```

```
# Train and evaluate each classifier
for classifier_name, classifier in classifiers:
```

```
classifier.fit(X_train, y_train)
y_pred = classifier.predict(X_test)

accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred)

print(f"Classifier: {classifier_name}")
print(f"Accuracy: {accuracy:.2f}")
print(f"Precision: {precision:.2f}")
print(f"Recall: {recall:.2f}")
print(f"F1-Score: {f1:.2f}")
print("\n")
```

```
Classifier: Support Vector Machine
Accuracy: 0.99
Precision: 1.00
Recall: 0.99
F1-Score: 0.99
```

```
Classifier: Logistic Regression
Accuracy: 0.98
Precision: 0.99
Recall: 0.98
F1-Score: 0.99
```

```
Classifier: Multinomial Naive Bayes
Accuracy: 0.93
Precision: 0.93
Recall: 0.93
F1-Score: 0.93
```

```
Classifier: Random Forest
Accuracy: 1.00
Precision: 1.00
Recall: 1.00
F1-Score: 1.00
```