Fake News Detection Using NLP - Brief Source Code and Output

Source Code:

```
import pandas as pd
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 classification_report
import nltk
from nltk.corpus import stopwords
import string

nltk.download('stopwords')
stop_words = set(stopwords.words('english'))

df = pd.read_csv('fake_or_real_news.csv')

def clean_text(text):
    text = text.lower()
    text = ''.join([char for char in text if char not in string.punctuation])
    tokens = text.split()
    tokens = [word for word in tokens if word not in stop_words]
    return ' '.join(tokens)

df['text'] = df['text'].apply(clean_text)

tfidf = TfidfVectorizer(max_features=5000)
X = tfidf.fit_transform(df['text'])
y = df['label'].apply(lambda x: 1 if x == 'FAKE' else 0)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = LogisticRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))
```

Sample Output (Classification Report):

```
precision    recall  f1-score   support

           0       0.98      0.97      0.98      1034
           1       0.96      0.98      0.97       980

    accuracy                           0.97      2014
   macro avg       0.97      0.97      0.97      2014
weighted avg       0.97      0.97      0.97      2014
```