```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
import seaborn as sns
```

```
df=pd.read_csv('news.csv')
df
```

| | Unnamed: 0 | title | text | label |
|---|---|---|---|---|
| 0 | 8476 | You Can Smell Hillary's Fear | Daniel Greenfield, a Shillman Journalism Fello... | FAKE |
| 1 | 10294 | Watch The Exact Moment Paul Ryan Committed Pol... | Google Pinterest Digg Linkedin Reddit Stumbleu... | FAKE |
| 2 | 3608 | Kerry to go to Paris in gesture of sympathy | U.S. Secretary of State John F. Kerry said Mon... | REAL |
| 3 | 10142 | Bernie supporters on Twitter erupt in anger ag... | — Kaydee King (@KaydeeKing) November 9, 2016 T... | FAKE |
| 4 | 875 | The Battle of New York: Why This Primary Matters | It's primary day in New York and front-runners... | REAL |
| ... | ... | ... | ... | ... |
| 6330 | 4490 | State Department says it can't find emails fro... | The State Department told the Republican Natio... | REAL |
| 6331 | 8062 | The 'P' in PBS Should Stand for 'Plutocratic' ... | The 'P' in PBS Should Stand for 'Plutocratic' ... | FAKE |
| 6332 | 8622 | Anti-Trump Protesters Are Tools of the Oligarc... | Anti-Trump Protesters Are Tools of the Oligar... | FAKE |
| 6333 | 4021 | In Ethiopia, Obama seeks progress on peace, se... | ADDIS ABABA, Ethiopia —President Obama convene... | REAL |
| 6334 | 4330 | Jeb Bush Is Suddenly Attacking Trump. Here's W... | Jeb Bush Is Suddenly Attacking Trump. Here's W... | REAL |

6335 rows × 4 columns

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.size
```

```
25340
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6335 entries, 0 to 6334
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   Unnamed: 0  6335 non-null   int64 
 1   title       6335 non-null   object
 2   text        6335 non-null   object
 3   label       6335 non-null   object
dtypes: int64(1), object(3)
memory usage: 198.1+ KB
```

```
df.head()
```

| | Unnamed: 0 | title | text | label |
|---|---|---|---|---|
| 0 | 8476 | You Can Smell Hillary's Fear | Daniel Greenfield, a Shillman Journalism Fello... | FAKE |
| 1 | 10294 | Watch The Exact Moment Paul Ryan Committed Pol... | Google Pinterest Digg Linkedin Reddit Stumbleu... | FAKE |
| 2 | 3608 | Kerry to go to Paris in gesture of sympathy | U.S. Secretary of State John F. Kerry said Mon... | REAL |
| 3 | 10142 | Bernie supporters on Twitter erupt in anger ag... | — Kaydee King (@KaydeeKing) November 9, 2016 T... | FAKE |
| 4 | 875 | The Battle of New York: Why This Primary Matters | It's primary day in New York and front-runners... | REAL |

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df.tail()
```

| | Unnamed: 0 | title | text | label |
|---|---|---|---|---|
| **6330** | 4490 | State Department says it can't find emails fro... | The State Department told the Republican Natio... | REAL |
| **6331** | 8062 | The ‘P’ in PBS Should Stand for ‘Plutocratic’ ... | The ‘P’ in PBS Should Stand for ‘Plutocratic’ ... | FAKE |
| **6332** | 8622 | Anti-Trump Protesters Are Tools of the Oligarc... | Anti-Trump Protesters Are Tools of the Oligar... | FAKE |
| **6333** | 4021 | In Ethiopia, Obama seeks progress on peace, se... | ADDIS ABABA, Ethiopia —President Obama convene... | REAL |
| **6334** | 4330 | Jeb Bush Is Suddenly Attacking Trump. Here's W... | Jeb Bush Is Suddenly Attacking Trump. Here's W... | REAL |

```
df.describe()
```

| | Unnamed: 0 |
|---|---|
| **count** | 6335.000000 |
| **mean** | 5280.415627 |
| **std** | 3038.503953 |
| **min** | 2.000000 |
| **25%** | 2674.500000 |
| **50%** | 5271.000000 |
| **75%** | 7901.000000 |
| **max** | 10557.000000 |

```
df.shape
```

```
(6335, 4)
```

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

| | 0 |
|---|---|
| **Unnamed: 0** | 0 |
| **title** | 0 |
| **text** | 0 |
| **label** | 0 |

```
df.duplicated().sum()
```

```
0
```

```
df.dropna()
```

| | Unnamed: 0 | title | text | label |
|---|---|---|---|---|
| **0** | 8476 | You Can Smell Hillary's Fear | Daniel Greenfield, a Shillman Journalism Fello... | FAKE |
| **1** | 10294 | Watch The Exact Moment Paul Ryan Committed Pol... | Google Pinterest Digg Linkedin Reddit Stumbleu... | FAKE |
| **2** | 3608 | Kerry to go to Paris in gesture of sympathy | U.S. Secretary of State John F. Kerry said Mon... | REAL |
| **3** | 10142 | Bernie supporters on Twitter erupt in anger ag... | — Kaydee King (@KaydeeKing) November 9, 2016 T... | FAKE |
| **4** | 875 | The Battle of New York: Why This Primary Matters | It's primary day in New York and front-runners... | REAL |
| **...** | ... | ... | ... | ... |
| **6330** | 4490 | State Department says it can't find emails fro... | The State Department told the Republican Natio... | REAL |
| **6331** | 8062 | The 'P' in PBS Should Stand for 'Plutocratic' ... | The 'P' in PBS Should Stand for 'Plutocratic' ... | FAKE |
| **6332** | 8622 | Anti-Trump Protesters Are Tools of the Oligarc... | Anti-Trump Protesters Are Tools of the Oligar... | FAKE |
| **6333** | 4021 | In Ethiopia, Obama seeks progress on peace, se... | ADDIS ABABA, Ethiopia —President Obama convene... | REAL |
| **6334** | 4330 | Jeb Bush Is Suddenly Attacking Trump. Here's W... | Jeb Bush Is Suddenly Attacking Trump. Here's W... | REAL |

6335 rows × 4 columns

```
df.empty
```

```
False
```

```
target_column = df.columns[-1]
features = df.columns[:-1]
```

```
X = df[features]
y = df[target_column]
```

```
X = df[features]

categorical_features = ['title', 'text']

preprocessor = ColumnTransformer(
    transformers=[
        ('cat', OneHotEncoder(handle_unknown='ignore', sparse_output=False), categorical_features),
    ],
    remainder='passthrough'
)
```

```
X = preprocessor.fit_transform(X)
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
label_encoder = LabelEncoder()
y_train = label_encoder.fit_transform(y_train)
y_test = label_encoder.transform(y_test)
```

```
model = RandomForestClassifier(n_estimators=100, random_state=42)
```

```
model.fit(X_train, y_train)
```

```
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
y_pred = model.predict(X_test)
```

```
print("\nClassification Report:")
print(classification_report(y_test, y_pred))
print("\nAccuracy Score:", accuracy_score(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       628
           1       1.00      1.00      1.00       639

    accuracy                           1.00      1267
   macro avg       1.00      1.00      1.00      1267
weighted avg       1.00      1.00      1.00      1267
```

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)

plt.figure(figsize=(8, 6))
sns.heatmap(cm, annot=True, fmt="d", cmap="Blues",
            xticklabels=label_encoder.classes_, yticklabels=label_encoder.classes_)
plt.xlabel("Predicted")
plt.ylabel("True")
plt.title("Confusion Matrix")
plt.show()
```