

pip install pandas scikit-learn matplotlib seaborn

Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (2.2.2)

Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)

Requirement already satisfied: matplotlib in /usr/local/lib/python3.11/dist-packages (3.10.0)

Requirement already satisfied: seaborn in /usr/local/lib/python3.11/dist-packages (0.13.2)

Requirement already satisfied: numpy>=1.23.2 in /usr/local/lib/python3.11/dist-packages  
(from pandas) (2.0.2)

Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-  
packages (from pandas) (2.8.2)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages  
(from pandas) (2025.2)

Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages  
(from pandas) (2025.2)

Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from  
scikit-learn) (1.14.1)

Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages  
(from scikit-learn) (1.4.2)

Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-  
packages (from scikit-learn) (3.6.0)

Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages  
(from matplotlib) (1.3.2)

Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from  
matplotlib) (0.12.1)

Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages  
(from matplotlib) (4.57.0)

Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages  
(from matplotlib) (1.4.8)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages  
(from matplotlib) (24.2)

Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.11/dist-packages (from  
matplotlib) (11.1.0)

Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib) (3.2.3)

Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)

```
import pandas as pd
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
from sklearn.naive_bayes import MultinomialNB
```

```
from sklearn.svm import LinearSVC
```

```
from sklearn.metrics import accuracy_score, classification_report
```

```
# Load dataset
```

```
url = "https://raw.githubusercontent.com/justmarkham/pycon-2016-tutorial/master/data/sms.tsv"
```

```
df = pd.read_csv(url, sep='\t', header=None, names=['label', 'message'])
```

```
# Convert labels to binary (ham = 0, spam = 1)
```

```
df['label_num'] = df.label.map({'ham': 0, 'spam': 1})
```

```
# Split data
```

```
X_train, X_test, y_train, y_test = train_test_split(
    df['message'], df['label_num'], test_size=0.2, random_state=42
)
```

```
# Vectorize text
```

```
vectorizer = TfidfVectorizer()
```

```
X_train_vec = vectorizer.fit_transform(X_train)
```

```
X_test_vec = vectorizer.transform(X_test)
```

```

# Train Naive Bayes

nb_model = MultinomialNB()
nb_model.fit(X_train_vec, y_train)
nb_preds = nb_model.predict(X_test_vec)

# Train SVM

svm_model = LinearSVC()
svm_model.fit(X_train_vec, y_train)
svm_preds = svm_model.predict(X_test_vec)

# Print results

print("Naive Bayes Accuracy:", accuracy_score(y_test, nb_preds))
print(classification_report(y_test, nb_preds))

print("\nSVM Accuracy:", accuracy_score(y_test, svm_preds))
print(classification_report(y_test, svm_preds))

output:

```

Naive Bayes Accuracy: 0.9668161434977578

	precision	recall	f1-score	support
0	0.96	1.00	0.98	966
1	1.00	0.75	0.86	149
accuracy			0.97	1115
macro avg	0.98	0.88	0.92	1115
weighted avg	0.97	0.97	0.96	1115

SVM Accuracy: 0.9919282511210762

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.99	1.00	1.00	966
---	------	------	------	-----

1	0.99	0.95	0.97	149
---	------	------	------	-----

accuracy		0.99		1115
----------	--	------	--	------

macro avg	0.99	0.98	0.98	1115
-----------	------	------	------	------

weighted avg	0.99	0.99	0.99	1115
--------------	------	------	------	------