```
############################################################
#                                                          #
#                       EXPERIMENT 8                       #
#                    Naive Bayes Theorem                   #
#                  Nathan Cordeiro 22co09                  #
#                                                          #
############################################################

import pandas as pd
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report
from sklearn.utils import shuffle

# Load the larger dataset
data = pd.read_csv('../sample_text_data.csv', encoding='utf-8')

# Strip any extra spaces from the column names
data.columns = data.columns.str.strip()

# Shuffle the data
data = shuffle(data, random_state=42)

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

# Convert text data into numerical data using CountVectorizer
vectorizer = CountVectorizer()
X_train_vect = vectorizer.fit_transform(X_train)
X_test_vect = vectorizer.transform(X_test)

# Initialize and train the Naive Bayes classifier
nb_classifier = MultinomialNB()
nb_classifier.fit(X_train_vect, y_train)

# Predict the labels for the test set
y_pred = nb_classifier.predict(X_test_vect)

# Evaluate the classifier
accuracy = accuracy_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)
class_report = classification_report(y_test, y_pred, zero_division=0)
```

```
print(f'Accuracy: {accuracy}')
print('Confusion Matrix:')
print(conf_matrix)
print('Classification Report:')
print(class_report)
```

CSV FILE DATA:

```
text,label
This is a spam email. Click here to win a free iPhone.,spam
Great product highly recommended!,feedback
The customer service was terrible.,feedback
I love this new restaurant!,feedback
Beware of phishing scams!,spam
The food was amazing!,feedback
The delivery was late.,feedback
The website is difficult to navigate.,feedback
I'm very satisfied with the service.,feedback
The product is defective.,feedback
```


```
text,label
This is a spam email. Click here to win a free iPhone.,spam
Great product highly recommended!,feedback
The customer service was terrible.,feedback
I love this new restaurant!,feedback
Beware of phishing scams!,spam
The food was amazing!,feedback
The delivery was late.,feedback
The website is difficult to navigate.,feedback
I'm very satisfied with the service.,feedback
The product is defective.,feedback
```


OUTPUT:

1.

```
Accuracy: 1.0
Confusion Matrix:
[[2]]
Classification Report:
              precision    recall  f1-score   support

    feedback       1.00      1.00      1.00         2

    accuracy                           1.00         2
   macro avg       1.00      1.00      1.00         2
weighted avg       1.00      1.00      1.00         2
```

2.	Accuracy: 0.5 Confusion Matrix: [[0 1] [0 1]] Classification Report: precision recall f1-score support negative 0.00 0.00 0.00 1 positive 0.50 1.00 0.67 1 accuracy 0.50 2 macro avg 0.25 0.50 0.33 2 weighted avg 0.25 0.50 0.33 2