Student Name : Bimal Kandel

install or Impoprting the necessary libraries

```
 1. # Impoprting the necessary libraries

# Import Required Libraries
import pandas as pd
import numpy as np
import nltk
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 accuracy_score, precision_score,
recall_score, f1_score, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

import nltk
nltk.download('stopwords')      # Stopwords for preprocessing
nltk.download('punkt')          # Tokenizer for text processing
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/iambimalk/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to /Users/iambimalk/nltk_data...
[nltk_data]   Package punkt is already up-to-date!

True
```

Downloaded the Amazon Customer Reviews dataset from Kaggle and importing them

```
 3.#loading the csv file dataset
data = pd.read_csv("//Users/iambimalk/Downloads/7817_1.csv")

data
```

```
                        id       asins   brand  \
0     AVpe7AsMilAPnD_xQ78G  B00QJDU3KY  Amazon
1     AVpe7AsMilAPnD_xQ78G  B00QJDU3KY  Amazon
2     AVpe7AsMilAPnD_xQ78G  B00QJDU3KY  Amazon
3     AVpe7AsMilAPnD_xQ78G  B00QJDU3KY  Amazon
4     AVpe7AsMilAPnD_xQ78G  B00QJDU3KY  Amazon
...                    ...         ...     ...
1592  AVpfo9ukilAPnD_xfhuj  B00N08JJZW  Amazon
1593  AVpfo9ukilAPnD_xfhuj  B00N08JJZW  Amazon
1594  AVpfo9ukilAPnD_xfhuj  B00N08JJZW  Amazon
1595  AVpfo9ukilAPnD_xfhuj  B00N08JJZW  Amazon
```

```
1596  AVpfo9ukilAPnD_xfhuj  B00NO8JJZW  Amazon

                                                 categories colors  \
0                             Amazon Devices,mazon.co.uk    NaN
1                             Amazon Devices,mazon.co.uk    NaN
2                             Amazon Devices,mazon.co.uk    NaN
3                             Amazon Devices,mazon.co.uk    NaN
4                             Amazon Devices,mazon.co.uk    NaN
...                                                 ...    ...
1592  Amazon Devices & Accessories,Amazon Device Acc...    NaN
1593  Amazon Devices & Accessories,Amazon Device Acc...    NaN
1594  Amazon Devices & Accessories,Amazon Device Acc...    NaN
1595  Amazon Devices & Accessories,Amazon Device Acc...    NaN
1596  Amazon Devices & Accessories,Amazon Device Acc...    NaN

                 dateAddeded           dateUpdated
dimension  \
0     2016-03-08T20:21:53Z  2017-07-18T23:52:58Z  169 mm x 117 mm x
9.1 mm
1     2016-03-08T20:21:53Z  2017-07-18T23:52:58Z  169 mm x 117 mm x
9.1 mm
2     2016-03-08T20:21:53Z  2017-07-18T23:52:58Z  169 mm x 117 mm x
9.1 mm
3     2016-03-08T20:21:53Z  2017-07-18T23:52:58Z  169 mm x 117 mm x
9.1 mm
4     2016-03-08T20:21:53Z  2017-07-18T23:52:58Z  169 mm x 117 mm x
9.1 mm
...                    ...                   ...
...
1592  2016-04-02T14:40:43Z  2017-08-13T08:28:46Z
NaN
1593  2016-04-02T14:40:43Z  2017-08-13T08:28:46Z
NaN
1594  2016-04-02T14:40:43Z  2017-08-13T08:28:46Z
NaN
1595  2016-04-02T14:40:43Z  2017-08-13T08:28:46Z
NaN
1596  2016-04-02T14:40:43Z  2017-08-13T08:28:46Z
NaN

      ean                                             keys  ...  \
0     NaN                     kindlepaperwhite/b00qjdu3ky  ...
1     NaN                     kindlepaperwhite/b00qjdu3ky  ...
2     NaN                     kindlepaperwhite/b00qjdu3ky  ...
3     NaN                     kindlepaperwhite/b00qjdu3ky  ...
4     NaN                     kindlepaperwhite/b00qjdu3ky  ...
...   ...                                             ...  ...
1592  NaN  alexavoiceremoteforamazonfiretvfiretvstick/b00...  ...
1593  NaN  alexavoiceremoteforamazonfiretvfiretvstick/b00...  ...
1594  NaN  alexavoiceremoteforamazonfiretvfiretvstick/b00...  ...
```

```
1595  NaN  alexavoiceremoteforamazonfiretvfiretvstick/b00...  ...
1596  NaN  alexavoiceremoteforamazonfiretvfiretvstick/b00...  ...

     reviews.rating                                 reviews.sourceURLs  
\
0               5.0  https://www.amazon.com/Kindle-Paperwhite-High-...

1               5.0  https://www.amazon.com/Kindle-Paperwhite-High-...

2               4.0  https://www.amazon.com/Kindle-Paperwhite-High-...

3               5.0  https://www.amazon.com/Kindle-Paperwhite-High-...

4               5.0  https://www.amazon.com/Kindle-Paperwhite-High-...

...             ...                                                ...

1592            3.0  https://www.amazon.com/Alexa-Voice-Remote-Amaz...

1593            1.0  https://www.amazon.com/Alexa-Voice-Remote-Amaz...

1594            1.0  https://www.amazon.com/Alexa-Voice-Remote-Amaz...

1595            3.0  https://www.amazon.com/Alexa-Voice-Remote-Amaz...

1596            1.0  https://www.amazon.com/Alexa-Voice-Remote-Amaz...


                                           reviews.text  \
0     I initially had trouble deciding between the p...   
1     Allow me to preface this with a little history...   
2     I am enjoying it so far. Great for reading. Ha...   
3     I bought one of the first Paperwhites and have...   
4     I have to say upfront - I don't like coroporat...   
...                                                 ...   
1592  This is not the same remote that I got for my ...   
1593  I have had to change the batteries in this rem...   
1594  Remote did not activate, nor did it connect to...   
1595  It does the job but is super over priced. I fe...   
1596  I ordered this item to replace the one that no...   

                                          reviews.title 
reviews.userCity  \
0                        Paperwhite voyage, no regrets!              
NaN   
1                     One Simply Could Not Ask For More              
NaN   
2            Great for those that just want an e-reader              
NaN   
3                              Love / Hate relationship              
NaN   
```

```
4                                             I LOVE IT
NaN
...                                                 ...              .
..
1592  I would be disappointed with myself if i produ...
NaN
1593                        Battery draining remote!!!!
NaN
1594      replacing an even worse remote. Waste of time
NaN
1595                                         Overpriced
NaN
1596  I am sending all of this crap back to amazon a...
NaN

     reviews.userProvince    reviews.username  sizes upc     weight
0                     NaN          Cristina M    NaN NaN  205 grams
1                     NaN               Ricky    NaN NaN  205 grams
2                     NaN       Tedd Gardiner    NaN NaN  205 grams
3                     NaN              Dougal    NaN NaN  205 grams
4                     NaN  Miljan David Tanic    NaN NaN  205 grams
...                   ...                 ...    ...  ..        ...
1592                  NaN     GregAmandawith4    NaN NaN   4 ounces
1593                  NaN     Amazon Customer    NaN NaN   4 ounces
1594                  NaN     Amazon Customer    NaN NaN   4 ounces
1595                  NaN          Meg Ashley    NaN NaN   4 ounces
1596                  NaN             DIANE K    NaN NaN   4 ounces

[1597 rows x 27 columns]
```

Clean the data by removing missing values or duplicates

```python
4. #Data Cleaning
data.drop_duplicates(inplace=True)
data.dropna(subset=['reviews.text', 'reviews.rating'], inplace=True)
```

Preprocess the text with tokenization, lowercasing, and removing stop words using NLTK

```python
# 5. Preprocessing
def preprocess_text(text):
    tokens = word_tokenize(text.lower())  # Tokenize and lowercase
    stop_words = set(stopwords.words('english'))
    filtered_tokens = [word for word in tokens if word.isalpha() and
word not in stop_words]
    return ' '.join(filtered_tokens)

# Ensure the 'reviews.text' column exists
data = data.rename(columns=lambda x: x.strip())  # Strip column names
to avoid leading/trailing spaces
```

```
if 'reviews.text' in data.columns:
    data['cleaned_review'] =
data['reviews.text'].apply(preprocess_text)  # Update 'reviews.text'
to your column name
else:
    raise ValueError("Column 'reviews.text' is missing in the
dataset.")
```

Label sentiment categories as Positive, Negative, or Neutral based on review ratings

```
# 6. Label Sentiment
# Assume rating column has values from 1-5, where 1-2 = Negative, 3 =
Neutral, 4-5 = Positive
def label_sentiment(rating):
    if rating <= 2:
        return 'Negative'
    elif rating == 3:
        return 'Neutral'
    else:
        return 'Positive'

if 'reviews.rating' in data.columns:
    data['sentiment'] = data['reviews.rating'].apply(label_sentiment)
else:
    raise ValueError("Column 'reviews.rating' is missing in the
dataset.")
```

Convert text to numerical features using TF-IDF

```
# 7. Convert Text to Numerical Features
if not data['cleaned_review'].str.strip().replace('',
pd.NA).isna().all():
    vectorizer = TfidfVectorizer(max_features=5000)
    X = vectorizer.fit_transform(data['cleaned_review']).toarray()
    y = data['sentiment']
else:
    raise ValueError("Cleaned reviews result in an empty vocabulary.
Ensure valid data is present.")
```

Split the dataset into training and testing sets with train_test_split

```
# 8. Split Dataset
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
```

Train a classifier using Logistic Regression

```
# 9. Train Classifier
model = LogisticRegression(max_iter=1000)
model.fit(X_train, y_train)
```

```
LogisticRegression(max_iter=1000)
```

Evaluate the model's performance using metrics like accuracy, precision, recall, and F1-score

```
# 10. Evaluate Model
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average='weighted')
recall = recall_score(y_test, y_pred, average='weighted')
f1 = f1_score(y_test, y_pred, average='weighted')
conf_matrix = confusion_matrix(y_test, y_pred, labels=['Positive',
'Neutral', 'Negative'])
```

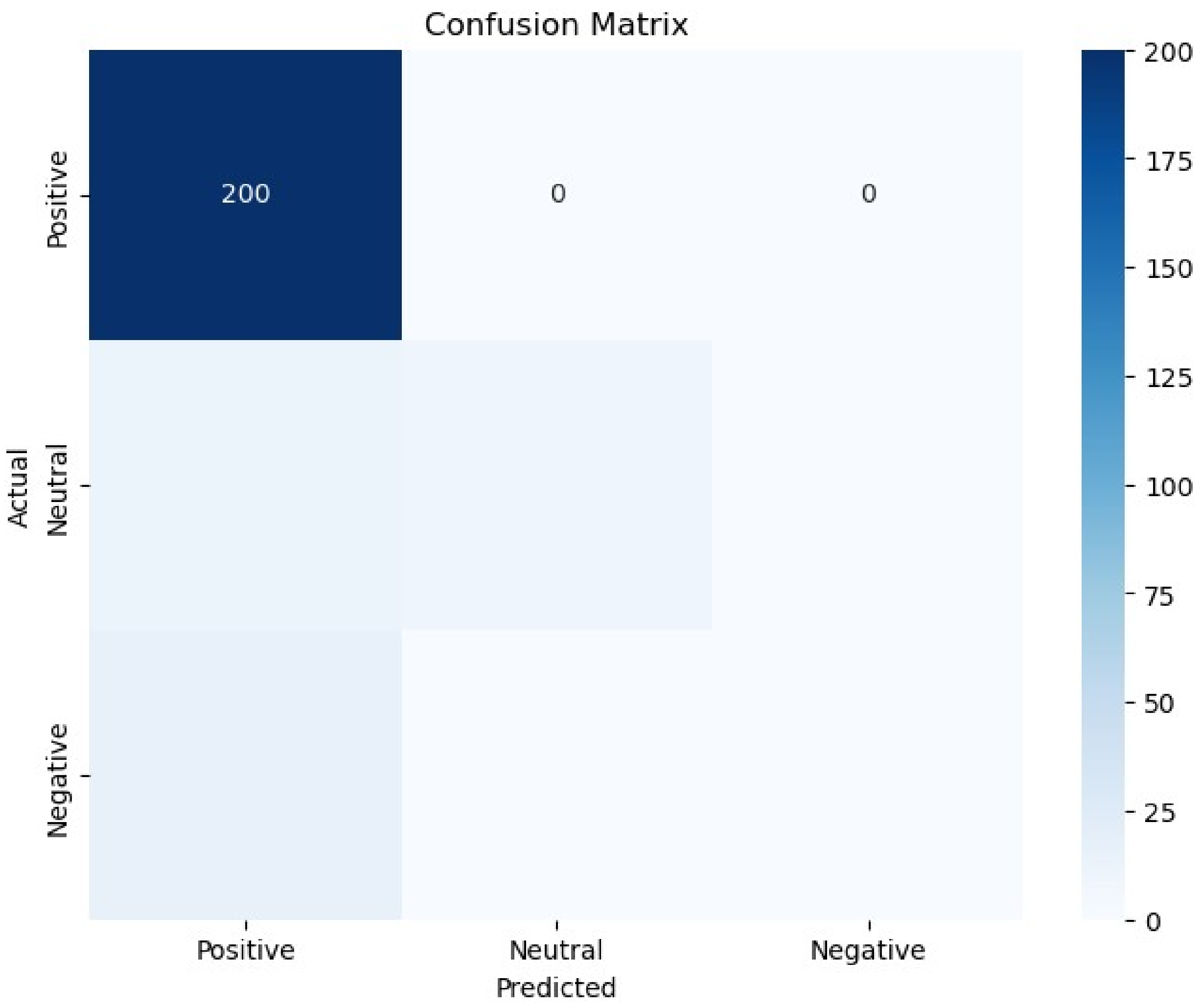
```
/opt/anaconda3/lib/python3.11/site-packages/sklearn/metrics/
_classification.py:1344: UndefinedMetricWarning: Precision is ill-
defined and being set to 0.0 in labels with no predicted samples. Use
`zero_division` parameter to control this behavior.
  _warn_prf(average, modifier, msg_start, len(result))
```

```
print(f"Accuracy: {accuracy}")
print(f"Precision: {precision}")
print(f"Recall: {recall}")
print(f"F1-Score: {f1}")
```

```
Accuracy: 0.8898305084745762
Precision: 0.8389455527223638
Recall: 0.8898305084745762
F1-Score: 0.8531432810458524
```

Visualize results with confusion matrices and sentiment distribution plots

```
# 11. Visualize Results
plt.figure(figsize=(8, 6))
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues',
xticklabels=['Positive', 'Neutral', 'Negative'],
yticklabels=['Positive', 'Neutral', 'Negative'])
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.title('Confusion Matrix')
plt.show()
```



```
# Sentiment distribution
if 'Sentiment' in data.columns:
    sentiment_counts = data['Sentiment'].value_counts()
    plt.figure(figsize=(6, 4))
    sentiment_counts.plot(kind='bar', color=['green', 'blue', 'red'])
    plt.title('Sentiment Distribution')
    plt.xlabel('Sentiment')
    plt.ylabel('Count')
    plt.show()
else:
    print("Column 'Sentiment' does not exist.")
```

```
Column 'Sentiment' does not exist.
```

```
14.# Save Results
data[['reviews.text', 'cleaned_review',
'sentiment']].to_csv('processed_reviews.csv', index=False)
```