Date: Ex No: 9	Title of the Lab Implementation of Natural language Processing for Amazon Customer Review Dataset	Name: Yuvraj Singh Chauhan Registration Number: RA1911027010058 Section: N1 Lab Batch: 1 Day Order: 3

Aim:

To find the Positive and Negative Reviews for Amazon Customer Review dataset.

Description of Amazon Customer Review:

The Internet has revolutionized the way we buy products. In the retail e-commerce world of online marketplace, where experiencing products are not feasible. Also, in today's retail marketing world, there are so many new products are emerging every day. Therefore, customers need to rely largely on product reviews to make up their minds for better decision making on purchase. However, searching and comparing text reviews can be frustrating for users. Hence we need better numerical ratings system based on the reviews which will make customers purchase decision with ease.

During their decision making process, consumers want to find useful reviews as quickly as possible using rating system. Therefore, models able to predict the user rating from the text review are critically important. Getting an overall sense of a textual review could in turn improve consumer experience. Also, it can help businesses to increase sales, and improve the product by understanding customer's needs.

Sentiment Analysis, also known as Opinion Mining, is one of the common research areas which performs Natural Language Processing (NLP) tasks for the purpose to extract subjective information by analyzing text data written by users. In the case of sentiment analysis of review data, the main goal is to identify the user's subjectivity and classify the statements into different groups of sentiments.

The amazon review dataset for electronics products were considered. The reviews given by the user to different products as well as reviews about user's experience with the product(s) were also considered. In this program, we aim to perform a sentiment analysis of product reviews written by online users from Amazon. This problem will be viewed as a multi-classification process and we seek to predict the sentiment scale of the user reviews based on machine learning classifiers.

Dataset:

	label	review
0	pos	Stuning even for the non-gamer: This sound tra...
1	pos	The best soundtrack ever to anything.: I'm rea...
2	pos	Amazing!: This soundtrack is my favorite music...
3	pos	Excellent Soundtrack: I truly like this soundt...
4	pos	Remember, Pull Your Jaw Off The Floor After He...
5	pos	an absolute masterpiece: I am quite sure any o...
6	neg	Buyer beware: This is a self-published book, a...
7	pos	Glorious story: I loved Whisper of the wicked ...
8	pos	A FIVE STAR BOOK: I just finished reading Whis...
9	pos	Whispers of the Wicked Saints: This was a easy...

Importing Files for NLP:

```python
import os
import re
import string
import spacy
import nltk
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from spacy import displacy
from wordcloud import WordCloud
from sklearn.model_selection import train_test_split
from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import LinearSVC
from sklearn.metrics import confusion_matrix,classification_report,plot_confusion_matrix,accuracy_score
from sklearn.metrics import plot_precision_recall_curve,plot_roc_curve
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from sklearn.linear_model import LogisticRegression
from collections import Counter
import warnings
nltk.download("vader_lexicon")
warnings.filterwarnings('ignore')
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
```

```python
url = 'https://raw.githubusercontent.com/yuvrajsinghchauhan/Amazon-Customer-Review-NLP-/main/amazon_reviews.tsv'
```

Sentiment Analyzer:

```
analysis_sa = SentimentIntensityAnalyzer()
```

```
random_text = 'the most exciting book i have ever read'
print(analysis_sa.polarity_scores(random_text))
```

```
{'neg': 0.0, 'neu': 0.632, 'pos': 0.368, 'compound': 0.5413}
```

```
random_text = 'This was the worst film to ever disgrace the screen.'
print(analysis_sa.polarity_scores(random_text))
```

```
{'neg': 0.477, 'neu': 0.523, 'pos': 0.0, 'compound': -0.8074}
```

```python
df['score']= df['review'].apply(lambda rew: analysis_sa.polarity_scores(rew))
df['compound']= df['score'].apply(lambda score: score['compound'])
df['predict']= df['compound'].apply(lambda value: 'pos' if value >=0 else 'neg')
```

```
df.head(10)
```

	label	review	score	compound	predict
0	pos	Stuning even for the non-gamer: This sound tra...	{'neg': 0.088, 'neu': 0.669, 'pos': 0.243, 'co...	0.9454	pos
1	pos	The best soundtrack ever to anything.: I'm rea...	{'neg': 0.018, 'neu': 0.837, 'pos': 0.145, 'co...	0.8957	pos
2	pos	Amazing!: This soundtrack is my favorite music...	{'neg': 0.04, 'neu': 0.692, 'pos': 0.268, 'com...	0.9858	pos
3	pos	Excellent Soundtrack: I truly like this soundt...	{'neg': 0.09, 'neu': 0.615, 'pos': 0.295, 'com...	0.9814	pos
4	pos	Remember, Pull Your Jaw Off The Floor After He...	{'neg': 0.0, 'neu': 0.746, 'pos': 0.254, 'comp...	0.9781	pos
5	pos	an absolute masterpiece: I am quite sure any o...	{'neg': 0.014, 'neu': 0.737, 'pos': 0.249, 'co...	0.9900	pos
6	neg	Buyer beware: This is a self-published book, a...	{'neg': 0.124, 'neu': 0.806, 'pos': 0.069, 'co...	-0.8744	neg
7	pos	Glorious story: I loved Whisper of the wicked ...	{'neg': 0.064, 'neu': 0.588, 'pos': 0.349, 'co...	0.9908	pos
8	pos	A FIVE STAR BOOK: I just finished reading Whis...	{'neg': 0.113, 'neu': 0.712, 'pos': 0.174, 'co...	0.8353	pos
9	pos	Whispers of the Wicked Saints: This was a easy...	{'neg': 0.033, 'neu': 0.777, 'pos': 0.19, 'com...	0.8196	pos

Predict Review:

```
review_predictor = PredictReview()
```

```
import nltk
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
True
```
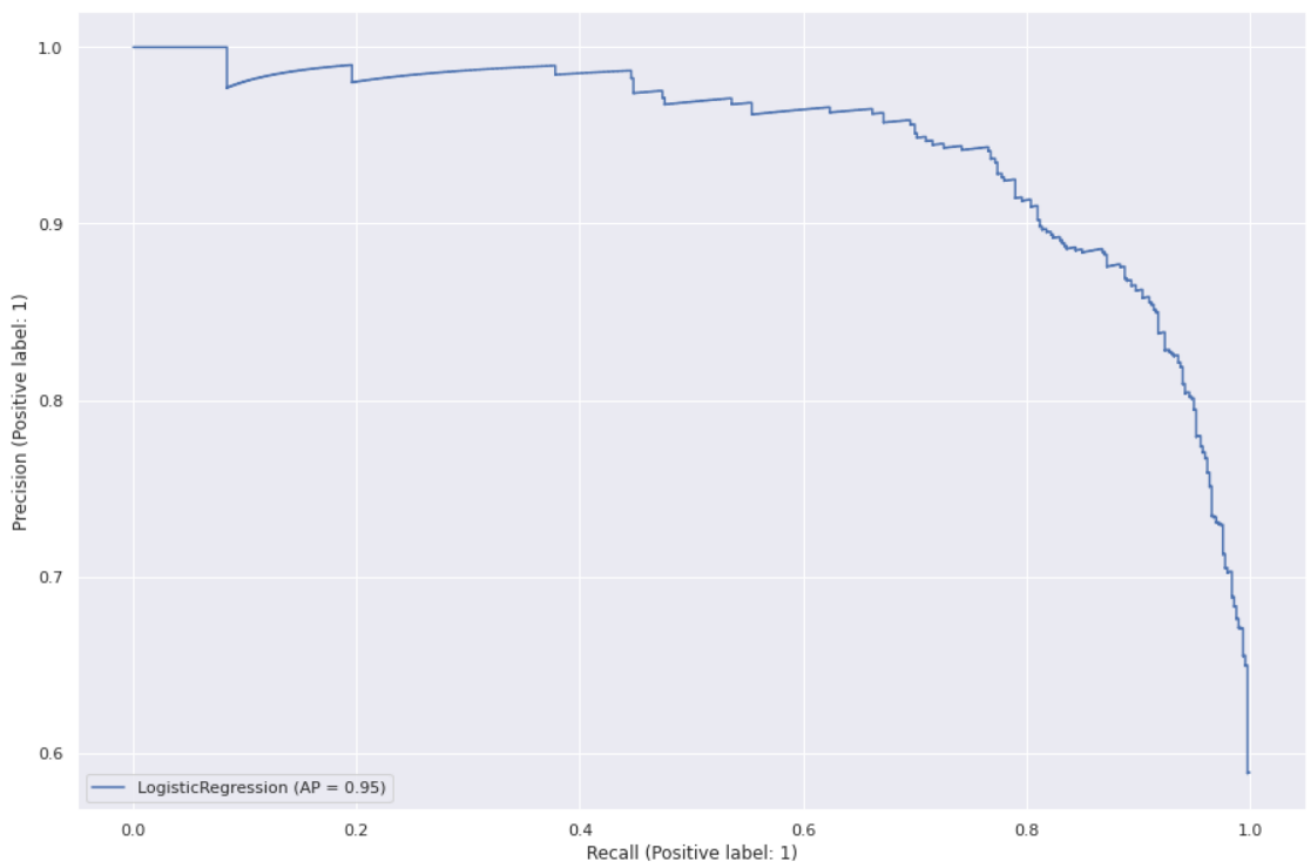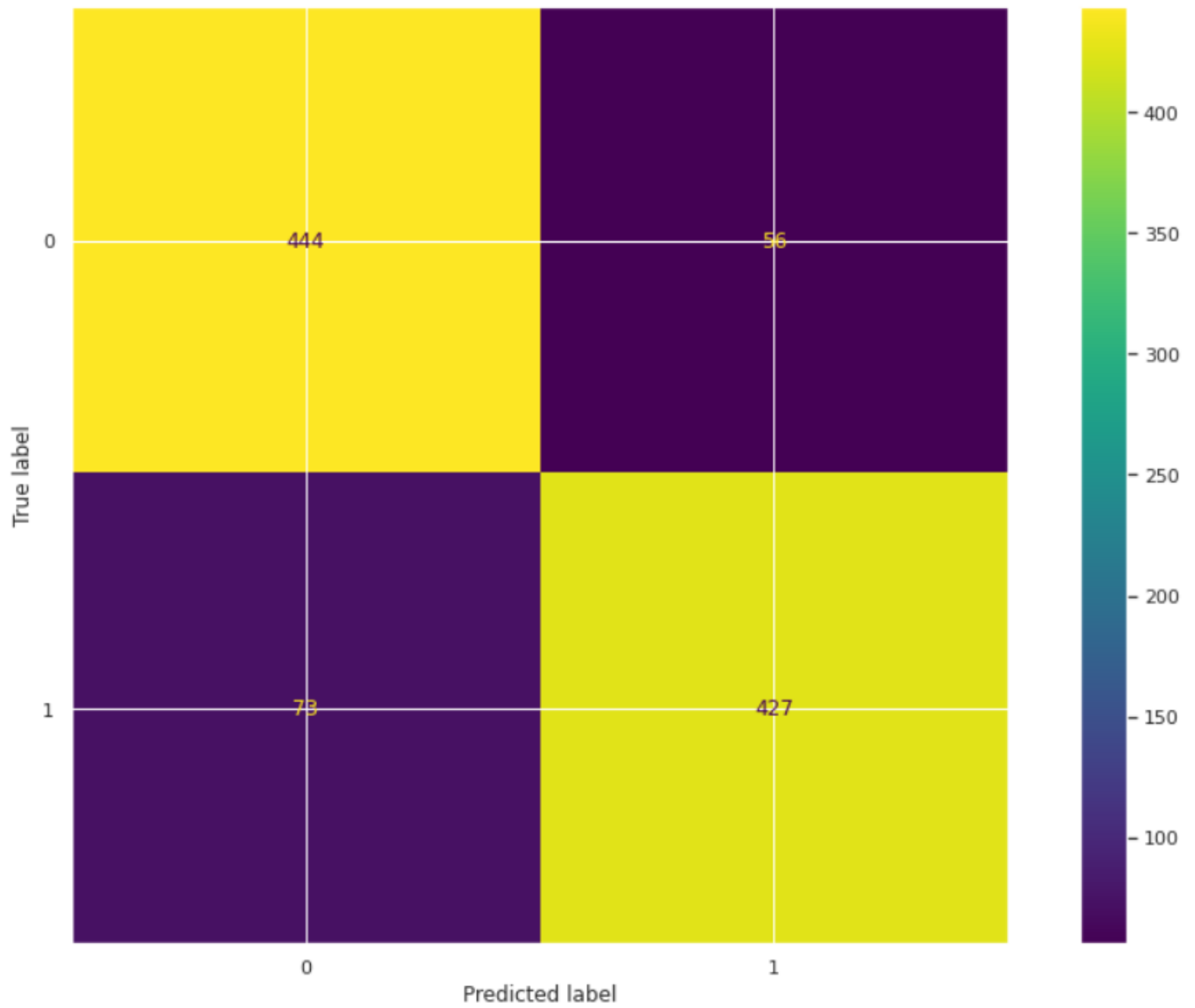
```
plt.rcParams["figure.figsize"] = (15,10)
model,coverter = review_predictor.base(df)
plt.show()
```
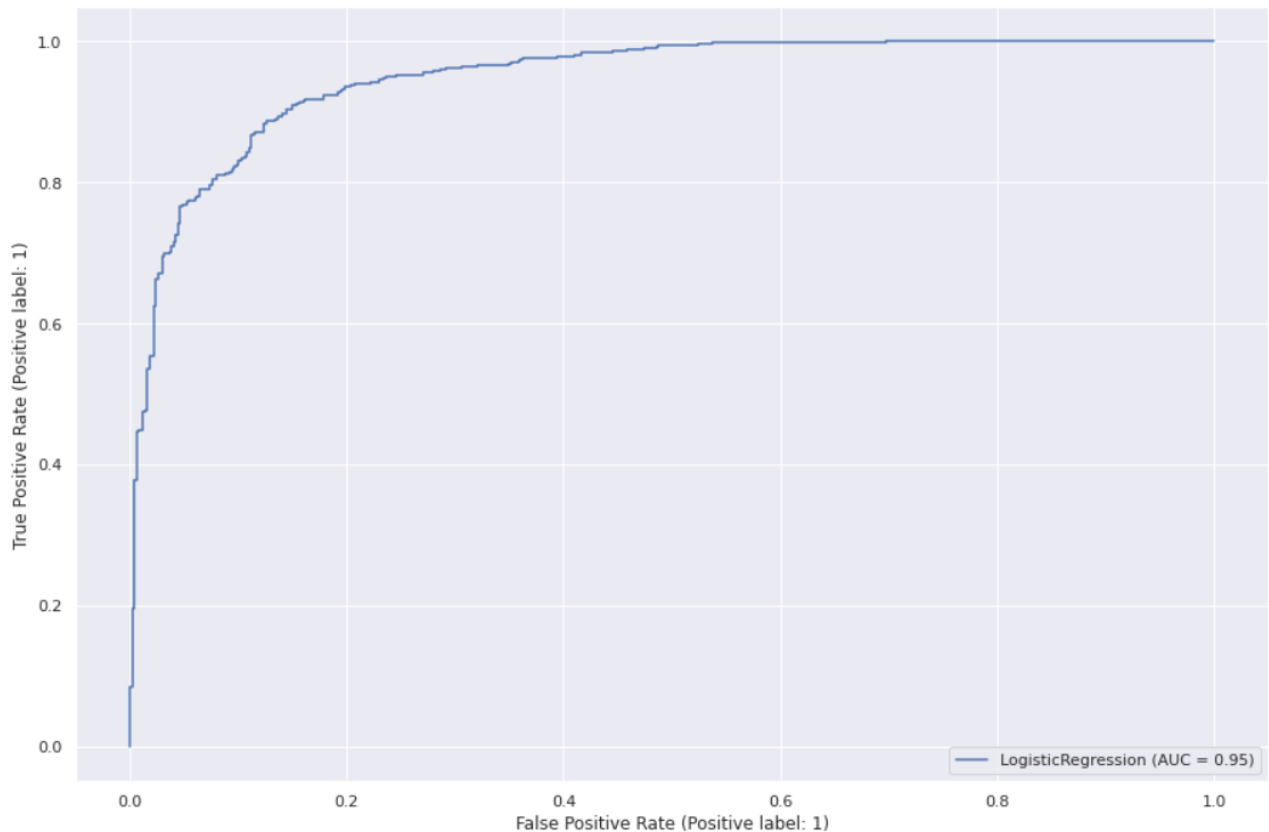
```
<sklearn.metrics._plot.confusion_matrix.ConfusionMatrixDisplay object at 0x7f3a9cc73ad0>
<sklearn.metrics._plot.precision_recall_curve.PrecisionRecallDisplay object at 0x7f3a9cc52150>
<sklearn.metrics._plot.roc_curve.RocCurveDisplay object at 0x7f3a9bc1fb90>
              precision    recall  f1-score   support

           0       0.89      0.86      0.87       517
           1       0.85      0.88      0.87       483

    accuracy                           0.87      1000
   macro avg       0.87      0.87      0.87      1000
weighted avg       0.87      0.87      0.87      1000
```

Overall accuracy is 87% 💣

```
<Figure size 864x864 with 0 Axes>
```

Quick Predictor:

```python
def quick_preictor(text):
    answer = review_predictor.test_sample(text,coverter,model)
    if answer == 'negative':
        print('\033[1m'+'\033[91m'+'Prediction is : '+answer+"\n")
        print(text)
    else:
        print('\033[1m'+'\033[92m'+'Prediction is : '+answer+"\n")
        print(text)
```

```python
text = 'Purchased this keyboard in September 2018 and it has already stopped working. Some keys do not work while others simply input a gibberish com
quick_preictor(text)
```

Prediction is : negative

Purchased this keyboard in September 2018 and it has already stopped working. Some keys do not work while others simply input a gibberish combination of characters. I also purchased the 4 year squaretrade warranty but since it is under the manufacturer warranty they wont allow me to make a claim.

```python
text = 'Love it just have to sync it up with all the other RGB but love it. Big Razor fan'
quick_preictor(text)
```

Prediction is : positive

Love it just have to sync it up with all the other RGB but love it. Big Razor fan

```python
text = "I get cold really easy and this jacket is great if you want a warm one. I'm 5'10 and weigh 135lbs and the small fits me well. It has fleece l
quick_preictor(text)
```

Prediction is : positive

I get cold really easy and this jacket is great if you want a warm one. I'm 5'10 and weigh 135lbs and the small fits me well. It has fleece lined pockets and a fleece lined hood. It has a good quality YKK zipper. The coat snaps shut in addition to zippering shut. It has elastic around the wrists for a nice tight fit that won't allow breezes to flow up your arm. There a fleece-line inside pocket about chest high big enough for a cell phone or wallet. It has a velcro enclosure.

```python
text = "Dont do it ! It will tell u steps and heartbeat .. But dont bother connecting it to your phone 👎 Not android anyways."
quick_preictor(text)
```

Prediction is : negative

Dont do it ! It will tell u steps and heartbeat .. But dont bother connecting it to your phone 👎 Not android anyways.

```
sns.set(rc={'figure.figsize':(5,5)})
sns.histplot(df, x="label", hue="label")
plt.show()
```
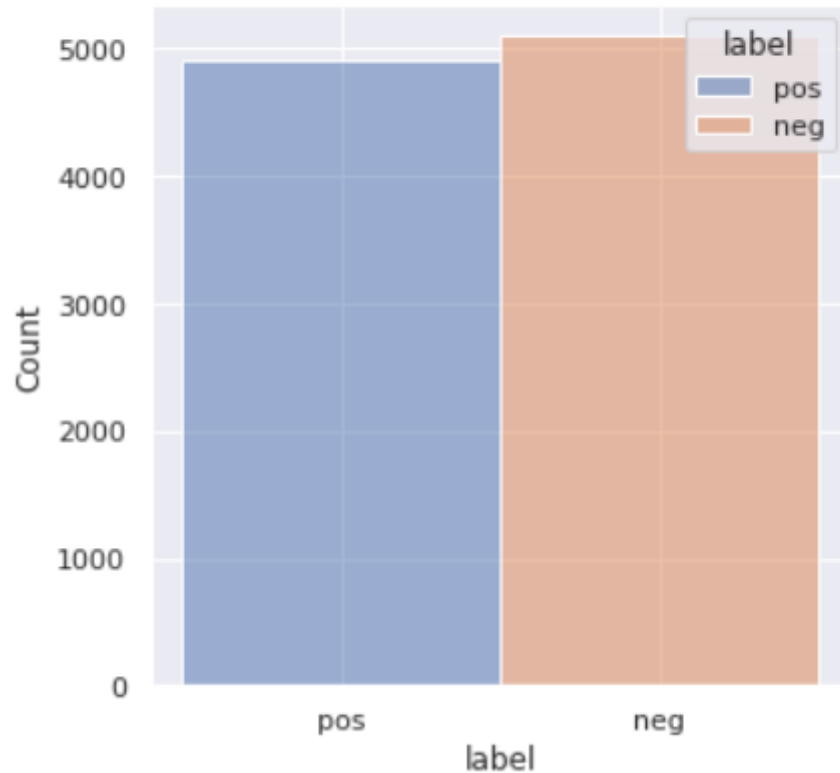

Conclusion:

Hence, Sentiment Analysis can successfully performed for Amazon Customer Review using NLP with an accuracy of 87%.

Signature of the Student

[YUVRAJ SINGH CHAUHAN]