

[Free Python 3 Tutorial](#) [Data Types](#) [Control Flow](#) [Functions](#) [List](#) [String](#) [Set](#) [Tuple](#) [Dictionary](#) [Oops](#)

Amazon Product Reviews Sentiment Analysis in Python

[Read](#)[Discuss](#)[Courses](#)[Practice](#)

Amazon gives a platform to small businesses and companies with modest resources to grow larger. And Because of its popularity, people actually spend time and write detailed reviews, about the brand and the product. So, by analyzing that data we can tell companies a lot about their products and also the ways to enhance the quality of the product. But that large amount of data can not be analyzed by a person.

Amazon Product Reviews Sentiment Analysis in Python

So here comes the Machine learning part, i.e. Natural Language Processing (NLP) to overcome the problem of large datasets and analyze it. Our task is to predict whether the review given is positive or negative. The real dataset after scraping the website might include millions of reviews. So we preprocessed the data for you,

Before starting the code, download the dataset by clicking the [link](#).

Steps to be followed

1. Importing Libraries and Datasets
2. Preprocessing and cleaning the reviews
3. Analysis of the Dataset
4. Converting text into Vectors
5. Model training, Evaluation, and Prediction

Let's start with the code now.

Importing Libraries and Datasets



We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

Got It !

- [Pandas](#) : For importing the dataset.
- [Scikit-learn](#) : For importing the model, accuracy module, and [TfidfVectorizer](#).
- [Warning](#) : To ignore all the warnings
- [Matplotlib](#) : To plot the visualization. Also used [Wordcloud](#) for that.

Python3

```
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
from sklearn.feature_extraction.text import TfidfVectorizer
import matplotlib.pyplot as plt
from wordcloud import WordCloud
```

For NLP part, we will be using [NLTK](#) Library. From that we will be requiring [stopword](#) and [punct](#). so let's download and import them using the below command.

Python3

```
import nltk
nltk.download('punct')
nltk.download('stopwords')
from nltk.corpus import stopwords
```

After that import the downloaded dataset using the below code.

Python3

```
data = pd.read_csv('AmazonReview.csv')
data.head()
```

Output :

	Review	Sentiment
0	Fast shipping but this product is very cheaply...	1
1	This case takes so long to ship and it's not e...	1
2	Good for not droids. Not good for iPhones. You...	1
3	The cable was not compatible between my macboo...	1
4	The case is nice but did not have a glow light...	1

Preprocessing and cleaning the reviews

Python3

```
data.info()
```

Output:

Data columns (total 2 columns):

```
#   Column      Non-Null Count  Dtype
---  -
0   Review      24999 non-null    object
1   Sentiment    25000 non-null    int64
```

Now, To drop the null values (if any), run the below command.

Python3

```
data.dropna(inplace=True)
```

To predict the Sentiment as positive(numerical value = 1) or negative(numerical value = 0), we need to change them the values to those categories. For that the condition will be like if the sentiment value is less than or equal to 3, then it is negative(0) else positive(1). For better understanding, refer the code below.

Python3

```
#4,5->positive(i.e 1)
data.loc[data['Sentiment']>3, 'Sentiment'] = 1
```

Now, once the dataset is ready, we will clean the review column by removing the stopwords. The code for that is given below.

Python3

```
stp_words=stopwords.words('english')
def clean_review(review):
    cleanreview=" ".join(word for word in review.
                          split() if word not in stp_words)
    return cleanreview

data['Review']=data['Review'].apply(clean_review)
```

Once we have done with the preprocess. Let's see the top 5 rows to see the improved dataset.

Python3

```
data.head()
```

Output :

	Review	Sentiment
0	Fast shipping product cheaply made I brought g...	0
1	This case takes long ship even worth DONT BUY!!!!	0
2	Good droids. Not good iPhones. You cannot use ...	0
3	The cable compatible macbook iphone. Also conn...	0
4	The case nice glow light. I'm disappointed pro...	0

Analysis of the Dataset

Let's check out that how many counts are there for positive and negative sentiments.

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).


```
plt.imshow(wordCloud.generate(consolidated), interpolation='bilinear')
plt.axis('off')
plt.show()
```

Output :



Now we have a clear picture of the words we have in both the categories.

Let's create the vectors.

Converting text into Vectors

TF-IDF calculates that how relevant a word in a series or corpus is to a text.

The meaning increases proportionally to the number of times in the text a word appears but is compensated by the word frequency in the corpus (data-set).

We will be implementing this with the code below.

Python3

```
cv = TfidfVectorizer(max_features=2500)
X = cv.fit_transform(data['Review']).toarray()
```

Model training, Evaluation, and Prediction

Once analysis and vectorization is done. We can now explore any machine learning model to train the data. But before that perform the train-test split.

הַיְּמִינִי

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy & Privacy Policy](#)

```
from sklearn.model_selection import train_test_split
x_train ,x_test,y_train,y_test=train_test_split(X,data['Sentiment'],
                                                test_size=0.25 ,
                                                random_state=42)
```

Now we can train any model, Let's explore the Logistic Regression.

Python3

```
from sklearn.linear_model import LogisticRegression

model=LogisticRegression()

#Model fitting
model.fit(x_train,y_train)

#testing the model
pred=model.predict(x_test)

#model accuracy
print(accuracy_score(y_test,pred))
```

Output :

0.81632

Let's see the [confusion matrix](#) for the results.

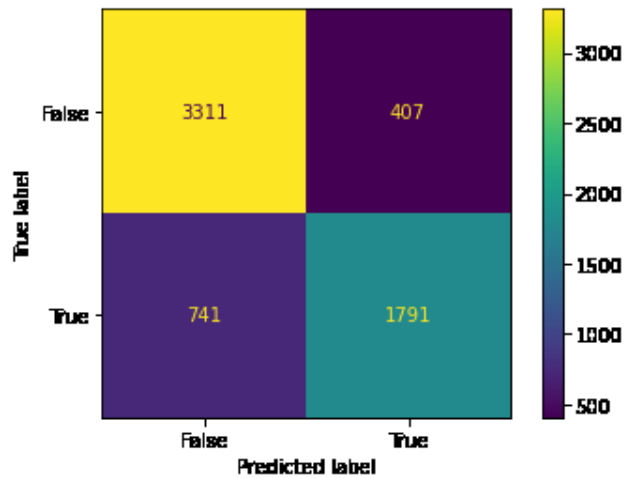
Python3

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

cm_display = metrics.ConfusionMatrixDisplay(confusion_matrix = cm,
                                             display_labels = [False, True])

cm_display.plot()
plt.show()
```

Output :



Whether you're preparing for your first job interview or aiming to upskill in this ever-evolving tech landscape, [GeeksforGeeks Courses](#) are your key to success. We provide top-quality content at affordable prices, all geared towards accelerating your growth in a time-bound manner. Join the millions we've already empowered, and we're here to do the same for you. Don't miss out - [check it out now!](#)

Last Updated : 21 Nov, 2022

4

[Previous](#)

[Next](#)

[Zillow Home Value \(Zestimate\) Prediction in ML](#)

[Music Recommendation System Using Machine Learning](#)

Similar Reads

[Flipkart Reviews Sentiment Analysis using Python](#)

[Amazon Product Review Sentiment Analysis using RNN](#)

[Python | NLP analysis of Restaurant reviews](#)

[Python | Sentiment Analysis using VADER](#)

[Facebook Sentiment Analysis using](#)

[Python | Sentiment Analysis using Affix](#)

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

[Sentiment Analysis of Hindi Text - Python](#)[Twitter Sentiment Analysis on Russia-Ukraine War Using Python](#)[Twitter Sentiment Analysis using Python](#)[Fine-tuning BERT model for Sentiment Analysis](#)

Complete Tutorials

[OpenAI Python API - Complete Guide](#)[Computer Vision Tutorial](#)[Pandas AI: The Generative AI Python Library](#)[Top Computer Vision Projects \(2023\)](#)[Python for Kids - Fun Tutorial to Learn Python Programming](#)

Article Contributed By :

D[deepakshi_mahajan](#)

deepakshi_mahajan

Vote for difficulty

Current difficulty : [Medium](#)[Easy](#)[Normal](#)[Medium](#)[Hard](#)[Expert](#)Improved By : [ruhelaa48](#)Article Tags : [Machine Learning Projects](#) , [Machine Learning](#) , [Python](#)Practice Tags : [Machine Learning](#), [python](#)

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).



A-143, 9th Floor, Sovereign Corporate
Tower, Sector-136, Noida, Uttar Pradesh -
201305



Company

About Us
Legal
Careers
In Media
Contact Us
Advertise with us
GFG Corporate Solution
Placement Training Program
Apply for Mentor

Languages

Python
Java
C++
PHP
GoLang
SQL
R Language
Android Tutorial

Data Science & ML

Explore

Job-A-Thon Hiring Challenge
Hack-A-Thon
GfG Weekly Contest
Offline Classes (Delhi/NCR)
DSA in JAVA/C++
Master System Design
Master CP
GeeksforGeeks Videos

DSA

Data Structures
Algorithms
DSA for Beginners
Basic DSA Problems
DSA Roadmap
Top 100 DSA Interview Problems
DSA Roadmap by Sandeep Jain
All Cheat Sheets

HTML & CSS

We use cookies to ensure you have the best browsing experience on our website. By using our site, you acknowledge that you have read and understood our [Cookie Policy](#) & [Privacy Policy](#).

[Data Science For Beginner](#)[CSS](#)[Machine Learning Tutorial](#)[Bootstrap](#)[Maths For Machine Learning](#)[Tailwind CSS](#)[Pandas Tutorial](#)[SASS](#)[NumPy Tutorial](#)[LESS](#)[Deep Learning Tutorial](#)[Web Design](#)

Computer Science

[GATE CS Notes](#)[Python Programming Examples](#)[Operating Systems](#)[Django Tutorial](#)[Computer Network](#)[Python Projects](#)[Database Management System](#)[Python Tkinter](#)[Software Engineering](#)[OpenCV Python Tutorial](#)[Digital Logic Design](#)[Python Interview Question](#)[Engineering Maths](#)

DevOps

[Git](#)[Top DS or Algo for CP](#)[AWS](#)[Top 50 Tree](#)[Docker](#)[Top 50 Graph](#)[Kubernetes](#)[Top 50 Array](#)[Azure](#)[Top 50 String](#)[GCP](#)[Top 50 DP](#)[DevOps Roadmap](#)[Top 15 Websites for CP](#)

System Design

[What is System Design](#)[TypeScript](#)[Monolithic and Distributed SD](#)[ReactJS](#)[High Level Design or HLD](#)[NextJS](#)[Low Level Design or LLD](#)[AngularJS](#)[Crack System Design Round](#)[NodeJS](#)[System Design Interview Questions](#)[Express.js](#)[Grokking Modern System Design](#)[Lodash](#)

Python

Competitive Programming

JavaScript

NCERT Solutions

NCERT Solutions for Class 12
NCERT Solution for Class 11
NCERT Solutions for Class 10
NCERT Solutions for Class 9
NCERT Solutions for Class 8
Complete Study Material

School Subjects

Mathematics
Physics
Chemistry
Biology
Social Science
English Grammar

Commerce

Accountancy
Business Studies
Indian Economics
Macroeconomics
Microeconomics
Statistics for Economics

Management & Finance

Management
HR Management
Income Tax
Finance
Economics

UPSC

Polity Notes
Geography Notes
History Notes
Science and Technology Notes
Economics Notes
Important Topics in Ethics
UPSC Previous Year Papers

SSC/ BANKING

SSC CGL Syllabus
SBI PO Syllabus
SBI Clerk Syllabus
IBPS PO Syllabus
IBPS Clerk Syllabus
SSC CGL Practice Papers

Colleges

Indian Colleges Admission & Campus Experiences
Top Engineering Colleges
Top BCA Colleges
Top MBA Colleges
Top Architecture College
Choose College For Graduation

Companies

IT Companies
Software Development Companies
Artificial Intelligence(AI) Companies
CyberSecurity Companies
Service Based Companies
Product Based Companies
PSUs for CS Engineers

[Company Wise Preparation](#)[JEE Mains](#)[Preparation for SDE](#)[JEE Advanced](#)[Experienced Interviews](#)[GATE CS](#)[Internship Interviews](#)[NEET](#)[Competitive Programming](#)[UGC NET](#)[Aptitude Preparation](#)[CAT](#)[Puzzles](#)

More Tutorials

[Software Testing](#)[Software Development](#)[Product Management](#)[SAP](#)[SEO](#)[Linux](#)[Excel](#)

Write & Earn

[Write an Article](#)[Improve an Article](#)[Pick Topics to Write](#)[Share your Experiences](#)[Internships](#)

@GeeksforGeeks, Sanchhaya Education Private Limited, All rights reserved