Sentiment Analysis Project Python:
The aim of the project is to scrap reviews from amazon, then analyse and come up with a machine learning (classifier) model that helps us predict the sentiments of the reviews of a product.

We have three files associated to the project:

1. Amazon_Reviews.py
2. Robo.py
3. NLPtry1.py

There are multiple stages involved in this project which is described below:

NOTE: Each of the programs above have been well documented to explain what they are doing
at each stage. This will be more of a summary.

Stage 1: Web Scraping:

Here the goal was to scrap reviews of a product from amazon. The product whose reviews were to be
scraped is user input to the program.

We used SELENIUM to automate the browsing process and SCRAPY(SPIDER) for web crawling and data extraction.

Robo.py is the selenium program that we use for automating the web browser
It is imported by the Amazon_Reviews.py class and is used to get links of the product page and review page

Amazon_Reviews.py is the scrapy program. It is used to extract all the reviews from the review pages. It is executed on the terminal so that the output thrown into the terminal is put into a file.

The data in the file(review.csv) is a table with two columns stars and reviews. It consists of about 1000 record which can vary based on the product.

Stage 2: Model Preparation

After we have extracted the data and stored it into a file, we now begin programming for our model.

All of this is present in NLPtry1.py

There are various stages in this too that is documented in the program.

To summarise it we have used natural language processing tools(NLP) to preprocess the data and make it ready for our model. We then create a sparse matrix of the words and the dependant variable y which is the scaled stars part that has been given 0 or 1 value.

We then use NaiveBayes model with a 20:80 split ratio and achieved a 84% accuracy rate which was determined based on the confusion matrix which is also visualised.