**Generic Sentiment Analysis**

**CSCE 5290 - Natural Language Processing**

**Professor - Dr. Sayed Khushal Shah**

**Project Increment - 1**

**Team Members:**

Vamshi Krishna Sharanammagari

Sai Krishna Reddy Beluri

Abhijit Talluri

Bharath Datta Chary Vadla

**Motivation:**

Usually, when we think of buying a mobile phone, we check many reviews through online websites like amazon. We can't read every review to know what are the positive, negative, and neutral reviews about the product. So, with the help of sentiment analysis, we can extract positive words, negative words, and neutral words. And we can analyze how to choose the best one based on reviews. Similarly, we can find out about the movies based on movie reviews.

**Significance:**

Sentiment analysis is a technique for assessing if a piece of text is good, negative, or neutral. It was essential for figuring out what other people were up to. Sentiment Analysis can reveal what people believe and how they act. A product review on the internet not only aids in analysis, but it also aids in marketing. the client, as well as the sector's personnel, to take the lead decision. This case study demonstrates how Opinion Mining can be used. This makes use of Natural Language Processing. Python is used to create the application. This paper includes the Sentiment. Feelings are divided into three categories: positive, neutral, and negative. It can be either positive or negative depending on the polarity.

**Objectives:**

With this idea, users can easily analyze the reviews and find out which product is suitable for their requirement. The goal is to extract expressions of opinion characterizing a target feature and classify it as positive or negative from a text including numerous features and varying perspectives.

**Features:**

There are three levels of emotion polarity categorization, depending on the breadth of the text: document level, sentence level, and entity and aspect level. Consider the following review: "I enjoy the multimedia functions, but the battery life is terrible." This statement evokes a range of emotions. The emotion is positive when it comes to multimedia, but bad when it comes to battery life. As a result, rather than considering the entire sentence and the general attitude, it is necessary to select only those opinions relevant to a certain feature (such as battery life or multimedia) and categorize them.

**Related Work:**

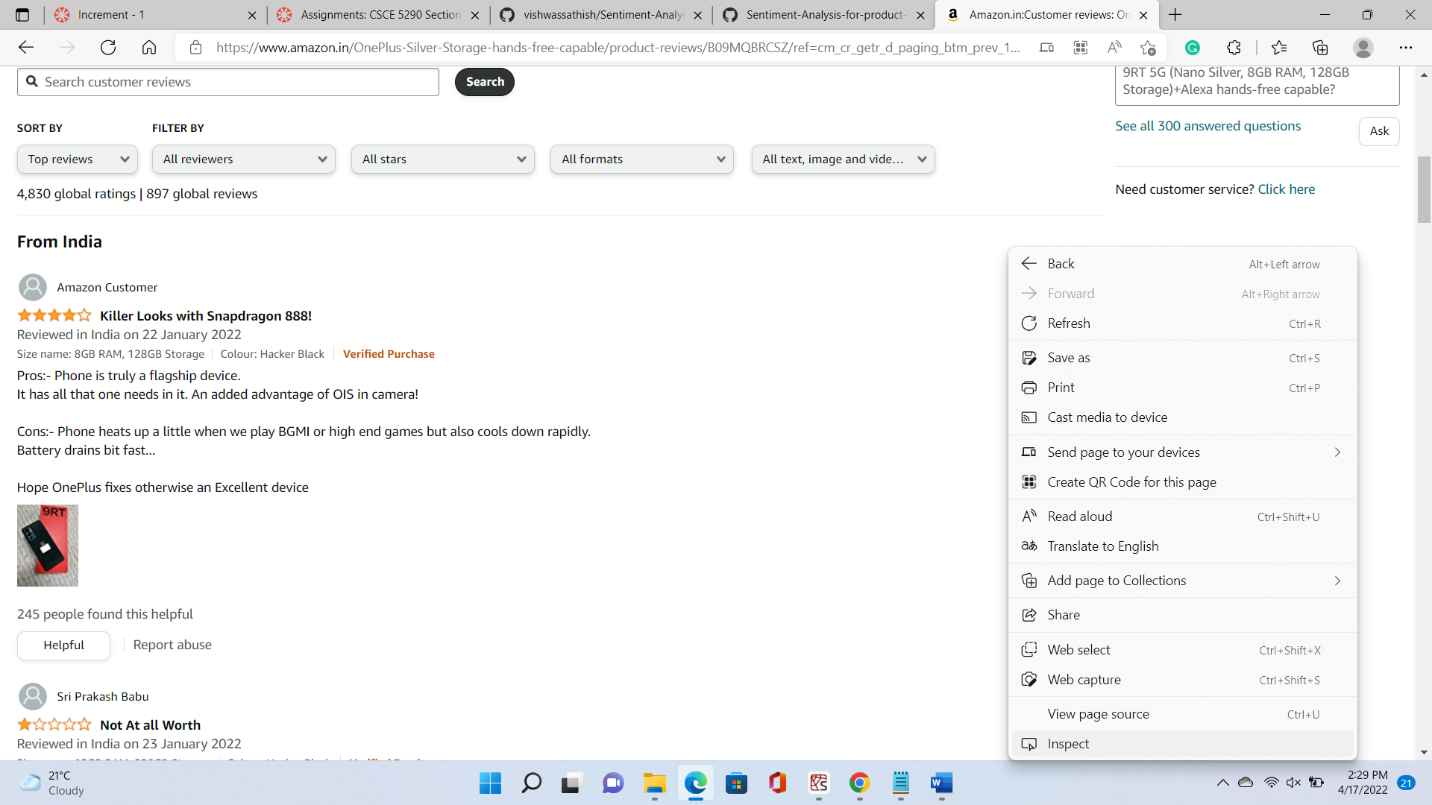
Vishwassathish uploaded his work (Sentiment-Analysis-for-product-reviews) on January 14, 2018. In his model he already had the dataset in a csv file format, which consists of two features, one is reviews and other feature is sentiment (positive or negative). It is direct in the dataset that the review is either positive or negative. After loading the dataset in the python, he converted the reviews into sentences and done tokenization using keras.preprocessing.text module. Later created training set and validation set by picking random reviews and by converting the reviews into list of tokens. Considering one list as one review. Thereafter created word2vec embeddings using GloVe to convert each word to of the review into 100-dimension tensor, which will be sent to the model for training. And used remaining for reviews for testing and finally printed the accuracy for predicting the review as positive or negative.

**Dataset:**

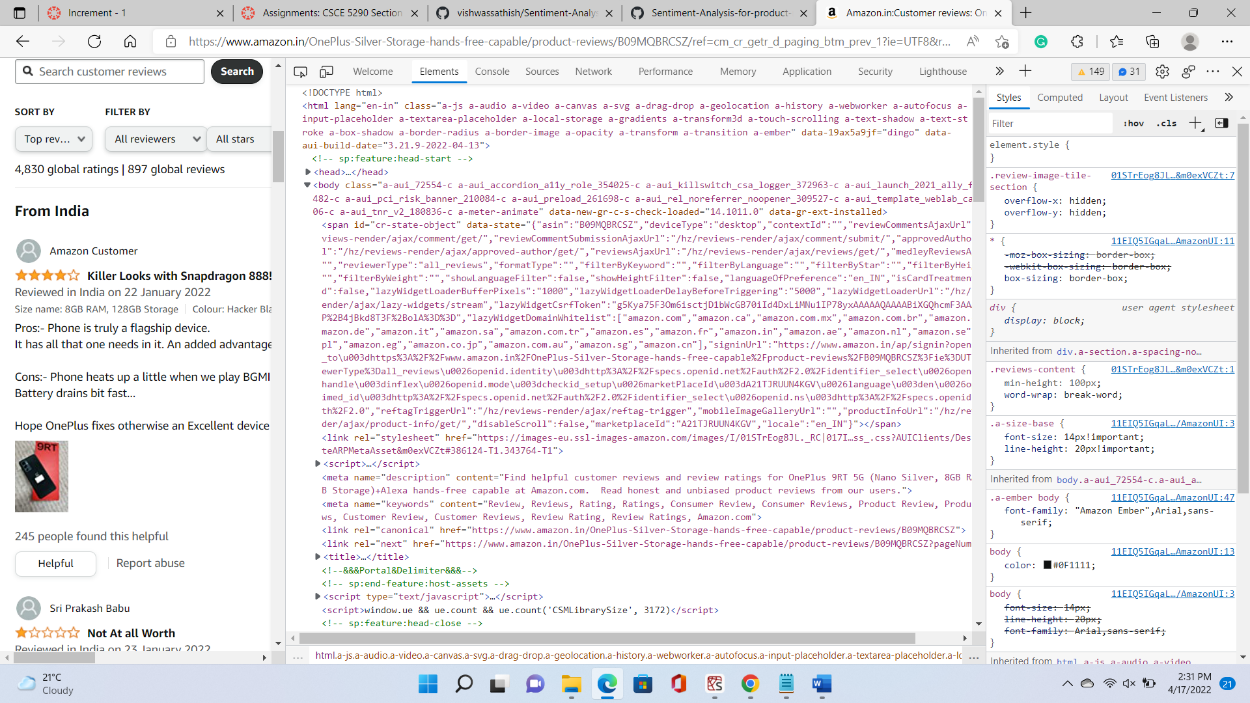
We used web scraping to build our dataset. i.e we collected data and content from the website of the product we want to analyze.

Steps:

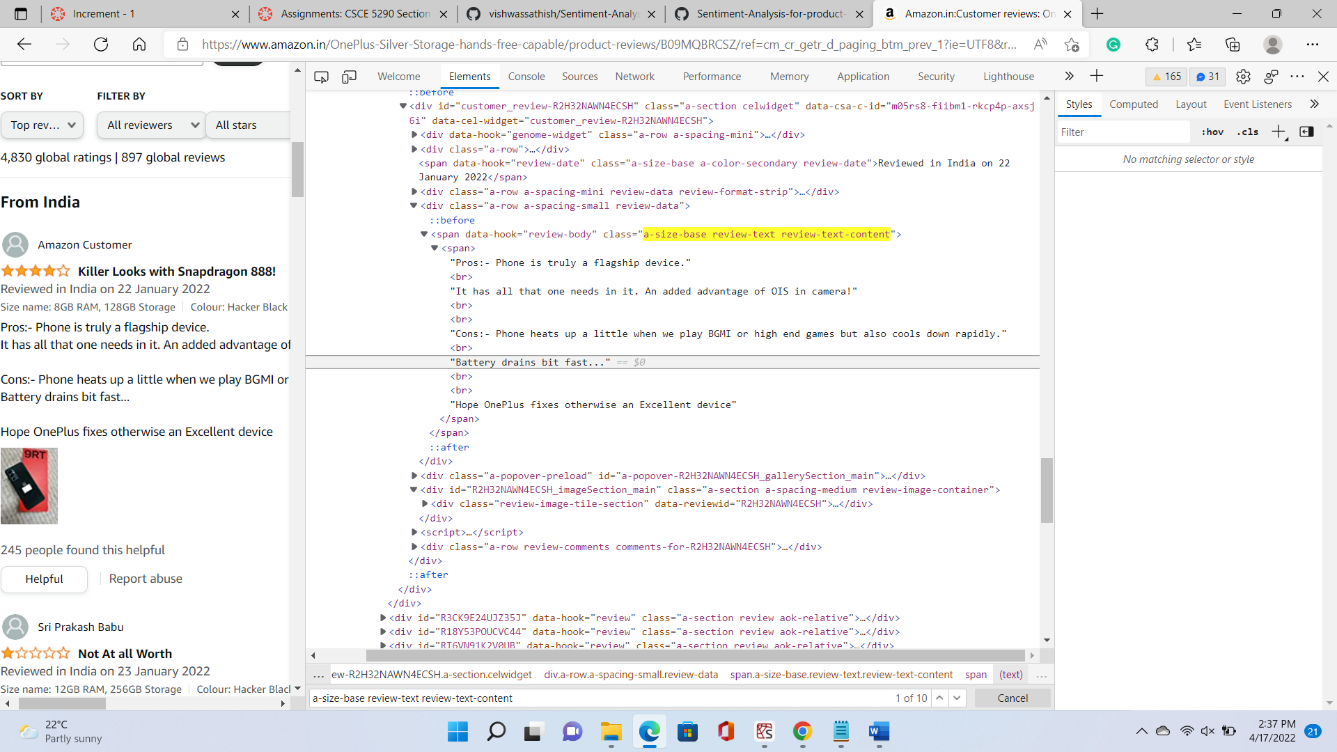
* [Amazon.in:Customer reviews: OnePlus 9RT 5G (Nano Silver, 8GB RAM, 128GB Storage)+Alexa hands-free capable](https://www.amazon.in/OnePlus-Silver-Storage-hands-free-capable/product-reviews/B09MQBRCSZ/ref=cm_cr_getr_d_paging_btm_prev_1?ie=UTF8&reviewerType=all_reviews&pageNumber1) , This is the website of the product we are extracting data from. After entering into that page, we just right click and select inspect from the options listed after right clicking the mouse as shown in below image:



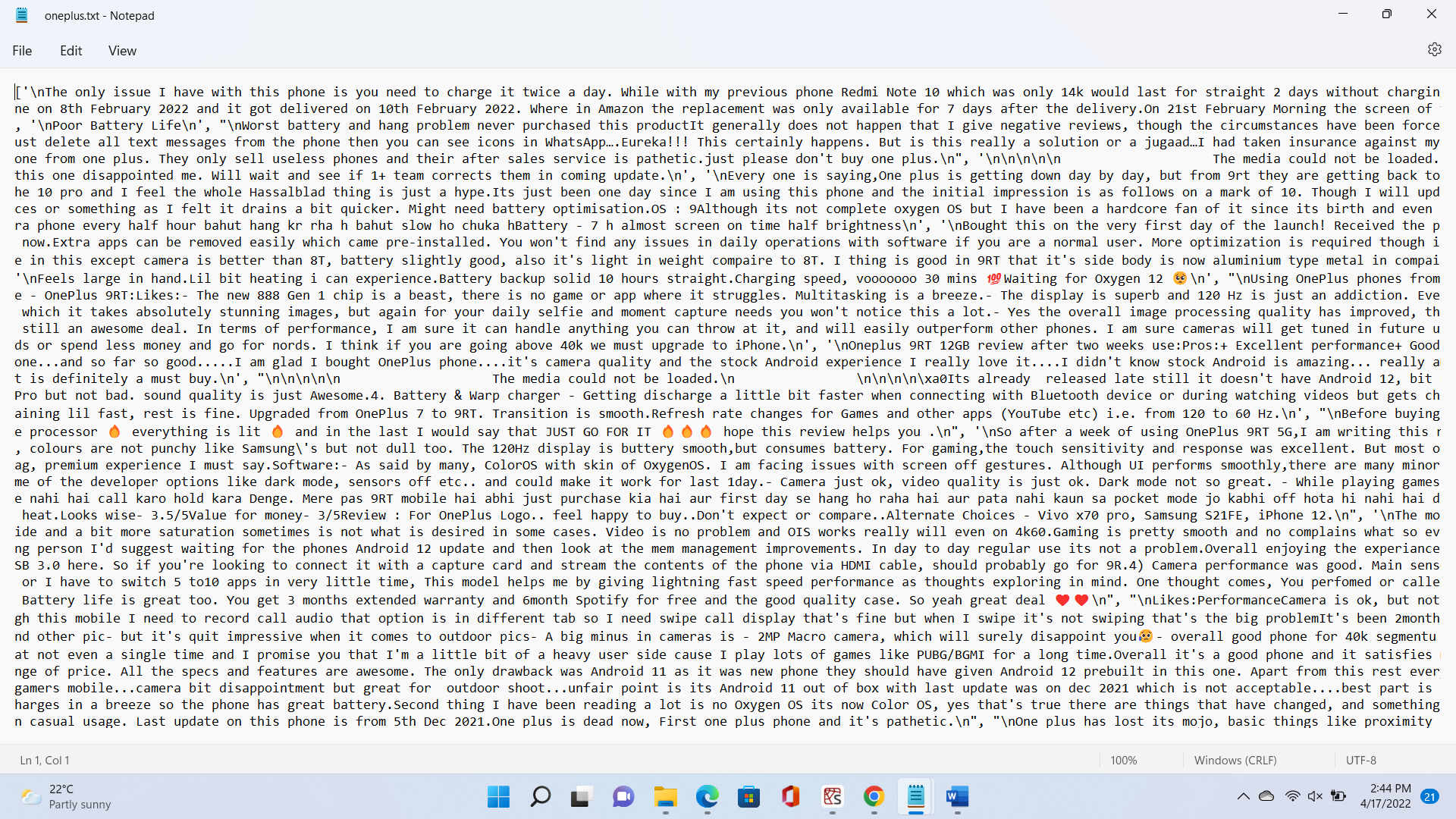
* And then html code of the webpage is displayed as shown in below image:



* Thereafter in html page we have tag called span and in that span tag we have class = “a-size-base review-text review-text-content”, where all the reviews are stored in that as shown in image below:

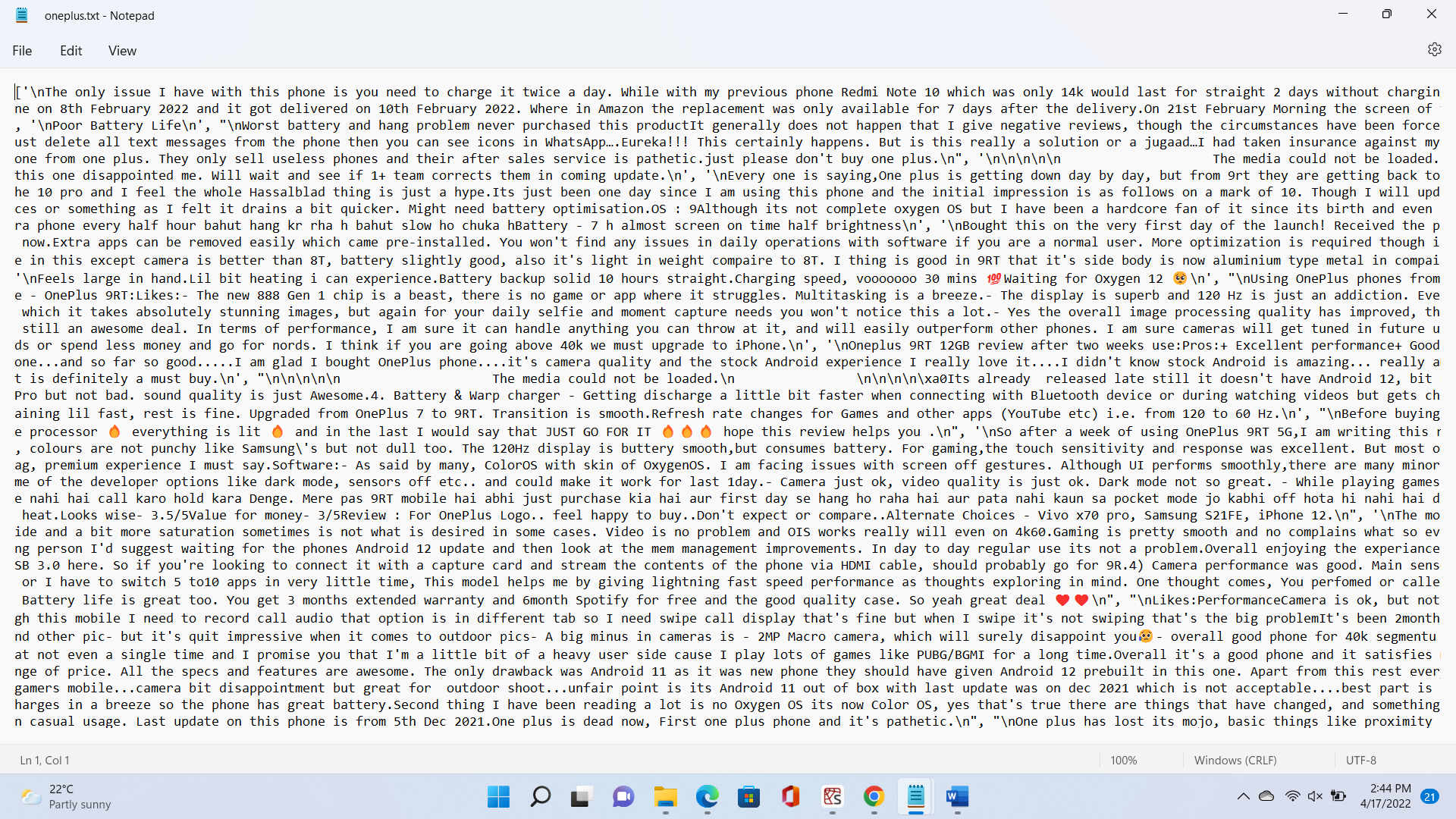


* And then all the reviews are stored in a list. Later joined all the reviews in a string format and then stored the data in txt file format (oneplus.txt).



**Detail Design of Features:**

Here in our project, we only have one feature in our dataset i.e reviews that all stored in a txt file. As shown in the below image:



**Analysis:**

Here in our project, we extract all the positive and negative words from the reviews on the product called oneplus which is a mobile. Based on these words we can tell what are the positive features of the mobile and what are the negative features of the mobile. With the help of all these words we can decide whether to buy that product or not. Same comes for all the products as we can’t read each and every review to find the positives and negatives of the product, we extract the words using Natural Language Processing techniques.

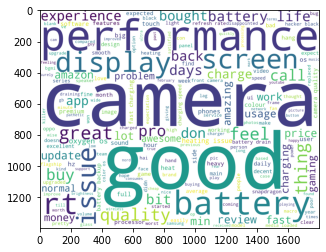
**Implementation:**

Initially we grab all the reviews of the product using its url and by using BeautifulSoup and requests modules and store them in a list. And then write all the reviews into a txt file. Later removing all the stop words, unnecessary symbols and most repeated words like oneplus, android, mobile etc. And then we have two other txt files pos.txt and neg.txt, using these two txt files we grab all the positive words and negative words, and store them in a positive and negative variables. And then showing all these words in a word cloud i.e positive words separate and negative words separate. Later using bigrams method from nltk module, we grab most repeated bigrams and display them in a word cloud. Here we have three word clouds positive words, negative words and bigrams. Here the most repeated words will be displayed with larger font and the font size keeps on decreasing based on repetition of a particular word.

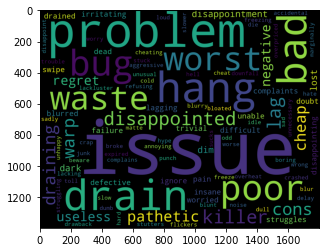
**Preliminary Results:**

Images of Word Clouds of positive words, negative words and most frequently occurred bigrams are shown below:

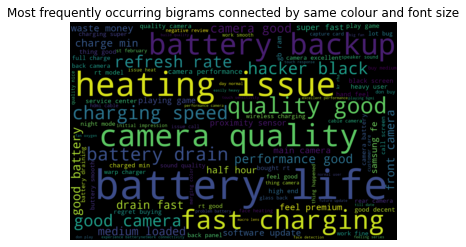
Positive Words:



Negative Words:



Most frequently repeated bigrams:



**Project Management:**

Implementation status report on sentiment analysis on product reviews:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.no | Name | Description | Responsibility | Contribution |
| 1. | Sharanammagari Vamshi Krishna | Coding and documentation | Coding and documentation | 25% |
| 2. | Sai Krishna Reddy Beluri | Coding and documentation | Documentation and coding | 25% |
| 3. | Abhijit Talluri | Coding and documentation | Coding and documentation | 25% |
| 4. | Bharath Datta Chary Vadla | Coding and documentation | Documentation and coding | 25% |

Work to be completed on sentiment analysis on movie reviews and advance models on sentiment analysis on product reviews:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.no | Name | Description | Responsibility | Contribution |
| 1. | Sharanammagari Vamshi Krishna | Coding and documentation | Documentation and coding | 25% |
| 2. | Sai Krishna Reddy Beluri | Coding and documentation | Coding and documentation | 25% |
| 3. | Abhijit Talluri | Coding and documentation | Documentation and coding | 25% |
| 4. | Bharath Datta Chary Vadla | Coding and documentation | Coding and documentation | 25% |

**References:**

1. [Sentiment Analysis-An Objective View | J4R - Journal for Research - Academia.edu](https://www.academia.edu/25824227/Sentiment_Analysis_An_Objective_View)
2. [IRJET-V6I9188.pdf](https://www.irjet.net/archives/V6/i9/IRJET-V6I9188.pdf)
3. [vishwassathish/Sentiment-Analysis-for-product-reviews: Sentiment Analysis using LSTM cells on Recurrent Networks. GloVe word embeddings were used for vector representation of words. Amazon Product Reviews were used as Dataset. (github.com)](https://github.com/vishwassathish/Sentiment-Analysis-for-product-reviews)