

**Python-Powered Sentiment Analysis Web App: Analyzing User Input and Datasets**

**Kaushik Parvathaneni, Simi Rajeev, Sadaramana Chowdam**

**Fall – 2023**

**Prof Olga Scrivner**

**Indiana University Bloomington**

**October 22, 2023.**

## **Project Description**

### **Objectives:**

Our objectives include design a web-based application to do sentiment analysis on any comments that user enters, it can be based on movie reviews, product reviews, or set of tweets related to any current affairs and also, we will allow users to enter text to analyze the polarity and subjectivity of the text. We plan to implement the functionality to clean the text, remove stop words and do some pre-processing. In addition, we plan to allow users to upload a set of tweets comments or reviews on csv/xlsx and calculate polarity and subjectivity scores and able to download the outcome.

### **Usefulness:**

Sentiment analysis is the process of identifying and extracting emotions from text data. Sentiment analysis, an important area in Natural Language Processing, is the process of automatically detecting affective states of text. It is widely applied to voice-of-customer materials such as product reviews in online shopping websites like Amazon, movie reviews or social media. It can be just a basic task of classifying the polarity of a text as being positive/negative or it can go beyond polarity, looking at sentiment states etc. Although there are other sentiment analysis tools, ours is more flexible. It can be used in a variety of fields, including sentiment analysis in social media as well as product evaluations and movie ratings.

Sentiment analysis refers to analyzing an opinion or feelings about something using data like text or images, regarding almost anything. Sentiment analysis helps companies in their decision-making process. For instance, if public sentiment towards a product is not so good, a company may try to modify the product or stop the production altogether to avoid any losses.

Our Sentiment Analysis Web App will be a flexible, user-friendly solution aimed at diverse user groups/stakeholders. Its distinctive qualities, which include real-time and batch analysis, customisation, and the capacity to address a wide range of applications, all make it a useful contribution to the sentiment analysis sector.

**Dataset:**

The dataset is obtained from Kaggle: <https://www.kaggle.com/datasets/tarkkaanko/amazon>

The dataset was compiled by the owner, Tark Kaan Koç, and is based on Amazon product data. The data appears to have been gathered in order to address significant e-commerce issues like the proper sorting of product comments and the calculation of after-sales product points. The data is presented in CSV format.

The dataset also includes a text field and other features such as reviewer names, product ratings, reviews, review times, days since evaluation, helpfulness indications, and votes. Yes, the data in this dataset are labelled.

The screenshot shows a Google Colab notebook with the following content:

```
[2] import pandas as pd
df=pd.read_csv('/content/amazon_reviews.csv')
```

```
[3] df
```

	Unnamed: 0	reviewerName	overall	reviewText	reviewTime	day_diff	helpful_yes	helpful_no	total_vote
0	0	NaN	4.0	No issues.	2014-07-23	138	0	0	0
1	1	Omle	5.0	Purchased this for my device, it worked as adv...	2013-10-25	409	0	0	0
2	2	1K3	4.0	it works as expected. I should have sprung for...	2012-12-23	715	0	0	0
3	3	1m2	5.0	This think has worked out great.Had a diff. br...	2013-11-21	382	0	0	0
4	4	2&amp;1/2Men	5.0	Bought it with Retail Packaging, arrived legit...	2013-07-13	513	0	0	0
...	...	...	...	...	...	...	...	...	...
4910	4910	ZM "J"	1.0	I bought this Sandisk 16GB Class 10 to use wit...	2013-07-23	503	0	0	0
4911	4911	Zo	5.0	Used this for extending the capabilities of my...	2013-08-22	473	0	0	0
4912	4912	Z S Liske	5.0	Great card that is very fast and reliable. It ...	2014-03-31	252	0	0	0
4913	4913	Z Taylor	5.0	Good amount of space for the stuff I want to d...	2013-09-16	448	0	0	0
4914	4914	Zza	5.0	I've heard bad things about this 64gb Micro SD...	2014-02-01	310	0	0	0

4915 rows x 12 columns

```
[4] df.isnull().sum()
```

Unnamed: 0	0
reviewerName	1
overall	0
reviewText	1
reviewTime	0
day_diff	0
helpful_yes	0
helpful_no	0

The dataset have 4915 rows and 12 columns and it have some null values as shown in the above figures. Data Cleaning is required here as it contains null values so we need to remove them or handle them in a way which will be used for us. We also need to preprocess the text and we may also need to normalize our data.

## **Functionalities:**

### Basic Functions:

The web app that we are proposing here will do sentiment analysis using TextBlob and it takes csv/xlsx files, analyze the sentiments in it and adds columns about sentiment scores of the review comments. It also features to enter review texts from users and do measure the polarity score and subjectivity score too. We will use streamlit as app framework for Machine Learning and Data Science and Pandas for data analysis and PyCharm IDE to implement the project. TextBlob from nltk library will be used for sentiment analysis.

### Advanced Functions:

If time permits, we will implement text summarization as well, that will help users to get a quicker consumption of large texts.

## **Communication and Sharing:**

We will have weekly meetings at **8 pm on every Monday using Zoom**. Our group files will be uploaded to github @ NLPD-590. We will collaborate in the Github folder <https://github.iu.edu/srajeed/NLPD-590>.

## **Milestones:**

Week 8 - project proposal (Group work finalized by Simi)

Week 10- NLP process (Group work will be finalized by Kaushik)

Week 12- Web app development (Group work will be finalized by Ramana)

Week 14- short video presentation (Group work will be finalized by All)

Week 15 - full demo (Group work to be finalized by ALL)

**Personal Contribution Statement:****Group Statement:**

I am happy to admit that each team member, including myself, contributed equally and significantly to the project idea that we worked on together. We all came to this initiative with the intention of dividing the work equally and preventing any one person from carrying an excessive burden.

**Individual Statement:**

Kaushik: I am a full time student so all I need to do is complete the course work and apply for jobs. I know each course have its own project or final exam but I can handle that and give priorities equally for all courses which I have been doing for past 6 years. So I am confident that I am able to complete all tasks or milestones in respective time period.

Simi: Based on our weekly meeting minutes, I will come up with the lists of tasks I am supposed to complete then I will figure out how much time I will have to spend for this project to prioritize it with my other schedules. I am planning to follow the SMART goal, using the right tools and maximizing the productivity, I am planning to achieve these milestones. Also I will be constantly connecting with my team mates to have a successful completion of the project.

Ramana: I will manage this project after my office hours and weekends.

**References:**

<https://www.kaggle.com/datasets/tarkkaanko/amazon>