**ASSIGNMENT XII**

**Project Title:**

Sentiment Analysis of Flipkart Reviews

**Team Leader:**

- Ejumalla Saikiran

**Team members:**

1. J. Sinduja

2. A. Bhavya Sri

3. S. Lakshanya

4. Naveen Rampa

**Submission Date:**

**07/04/2024**

**Methodology**

1. **DataSource**

The dataset is acquired from LMS platform.

The dataset consists of reviews of the products which customers purchased from Flipkart. The reviews given are the experience about the product customer bought through Flipkart and given a rating to that product. The dataset for Flipkart contains 3 columns:

Product\_name: the name of the product with little details about it.

Review: review what the customer gives about that product.

Rating: ranges from 1 to 5 stars given by the customer.

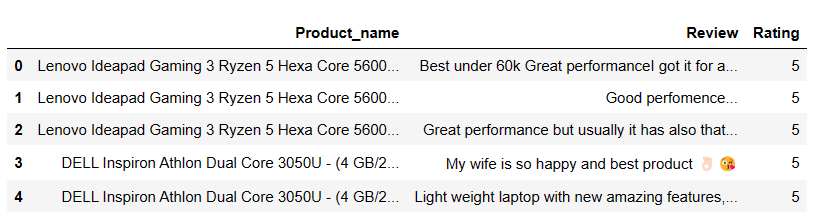


Fig 1.1 First 5 rows of the data

**Shape of the data**

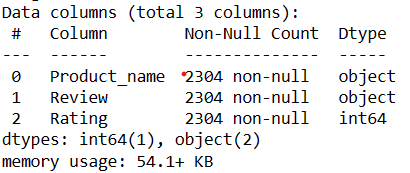
****

Fig 1.2 Shape of the data

The dataset comprises 2304 rows and 3 columns in total.

1. **Data Pre-processing**

**2.1 Analyzing Column Info:**

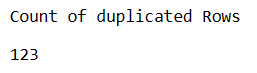


**Observations:**

* All three columns have 2304 non-null values, which means there are no missing entries in any of the columns.
* The data types of the columns are:
  + "Product\_name": object
  + "Review": object
  + "Rating": int64
* This tells us that the "Product\_name" and "Review" columns contain text data, while the "Rating" column contains integer data.

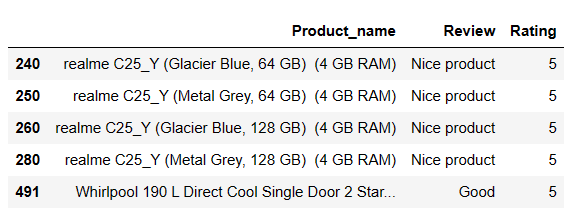
**2.2 Handling Duplicates:**

Duplicate rows are identical entries within a dataset that can skew analysis and should be removed for accuracy and data integrity.



Count of duplicated Rows

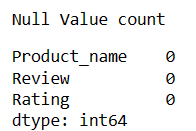
It is observed that a total of ‘123’ duplicated rows are present in the dataset.



Few duplicated nodes

After identifying duplicate rows in the dataset, they were removed to preserve data integrity and ensure accuracy in subsequent analyses.

**2.3 Handling Null values:**

****

The null value count is 0 for all three columns - "Product\_name", "Review", and "Rating". This means that there are no missing entries in any of the columns in this dataset.

1. **Sentiment Analysis Models Used**
   * VADER (Valence Aware Dictionary for Sentiment Reasoning):

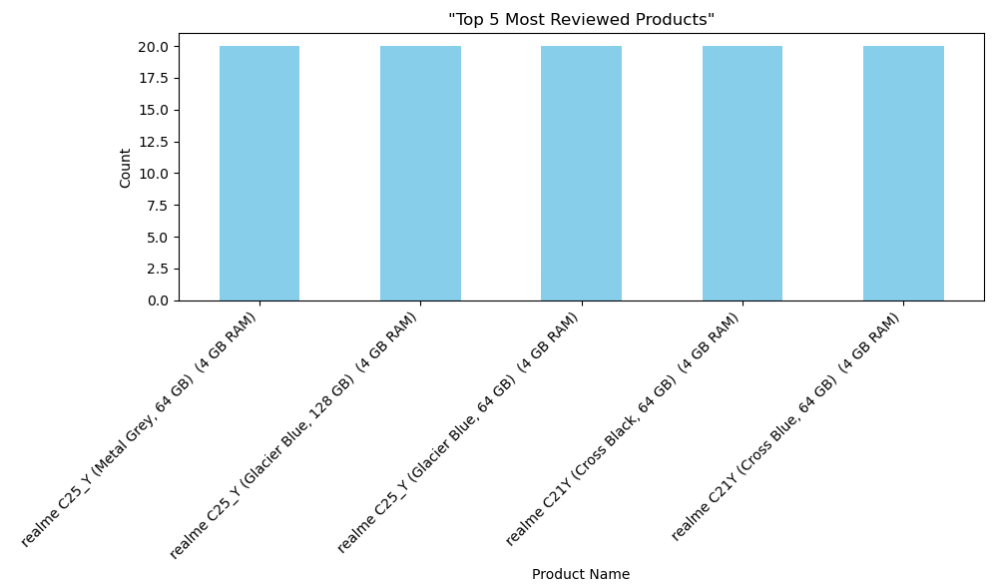
VADER is a lexicon and rule-based sentiment analysis tool designed for social media text. It provides sentiment scores for each input text, indicating the positivity, negativity, neutrality, and an overall compound score.

* + Roberta:

Roberta is a state-of-the-art deep learning model based on the Transformer architecture. It's trained on extensive text data and can capture complex linguistic patterns. In this analysis, a pre-trained Roberta model fine-tuned for sentiment analysis was used to classify text into positive, negative, or neutral categories, providing probabilities for each sentiment class.

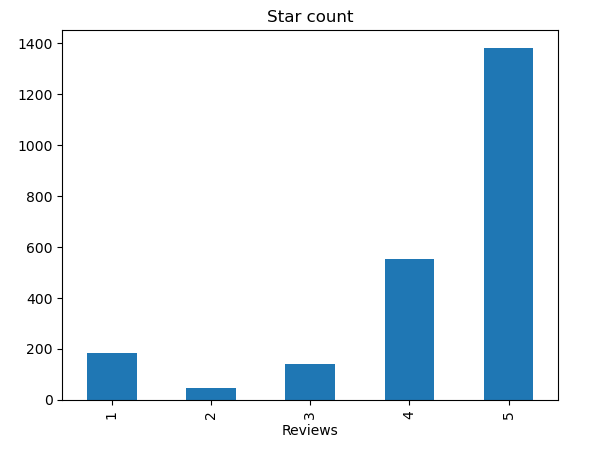
1. **Exploratory Data Analysis**

**4.1 Top 5 Most Reviewed Products**

****

**Observations:**

The top 5 most reviewed products include variants from the realme C25\_Y and realme C21Y smartphone models. Each variant, regardless of color or storage capacity, has received a consistent number of reviews, indicating widespread customer interest. The realme C25\_Y variants, such as Metal Grey and Glacier Blue options, have attracted significant attention, as have the realme C21Y variants in Cross Black and Cross Blue colors. These observations highlight strong engagement and positive reception from users toward these highly reviewed smartphone models.

****

**Observations:**

The graph displays the distribution of ratings across multiple reviewed products. Rating 5 is the most prevalent, indicating high levels of satisfaction among customers. Following closely is rating 4, suggesting overall positive feedback. However, there are also occurrences of lower ratings, including 1, 3, and 2, indicating areas for improvement or issues encountered by customers. Overall, the graph provides valuable insights into customer satisfaction levels across various products.

**Sentiment Analysis Results**

**VADER Sentiment Analysis**

The VADER model provides sentiment scores for each review in terms of negativity (neg), neutrality (neu), positivity (pos), and compound score.

The compound score ranges from -1 (extremely negative) to +1 (extremely positive), with 0 indicating a neutral sentiment.

1. Test on Negative Statement

2. Test on Neutral statement



3. Test on Positve Statement



The sentiment analysis results show a clear contrast among the provided statements. The first statement reflects a highly negative sentiment, expressing dissatisfaction with the product's quality and connectivity, as indicated by its compound score of -0.8748. In contrast, the second statement maintains a neutral tone, merely describing the product's appearance without strong emotion, with a compound score of 0.4215. Conversely, the third statement exudes strong positivity, expressing happiness and satisfaction with the product, with a compound score of 0.8808. These scores succinctly summarize the varying sentiments expressed, ranging from negative to neutral to positive.

**Roberta Sentiment Analysis**

1. Test on Negative Statement



2. Test on Neutral statement



3. Test on Positve Statement

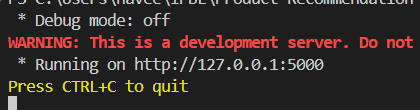


1. **Final Implementation**

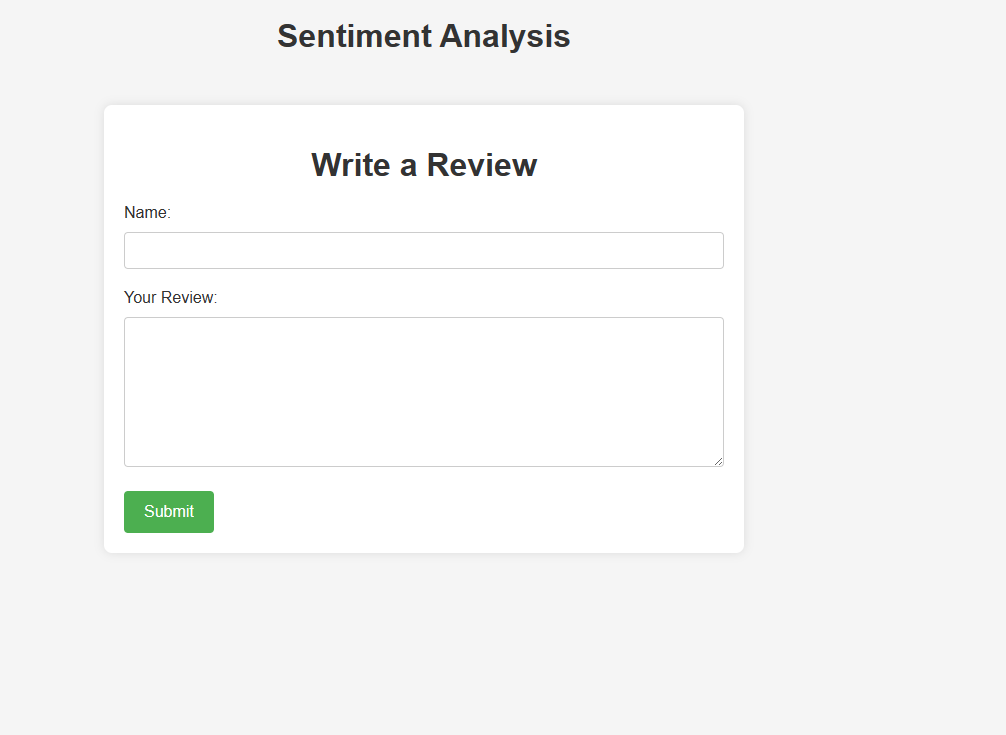
For the final implementation, a web application has been developed using Flask, demonstrating the implementation of sentiment analysis.

The demonstration of the website implementation is as follows:

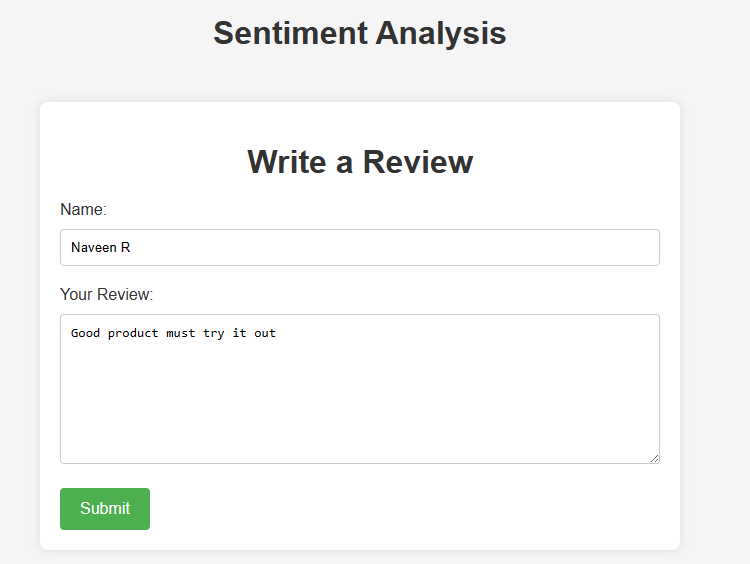
1. Launch the web app by executing the Python code, which will deploy the application on localhost.



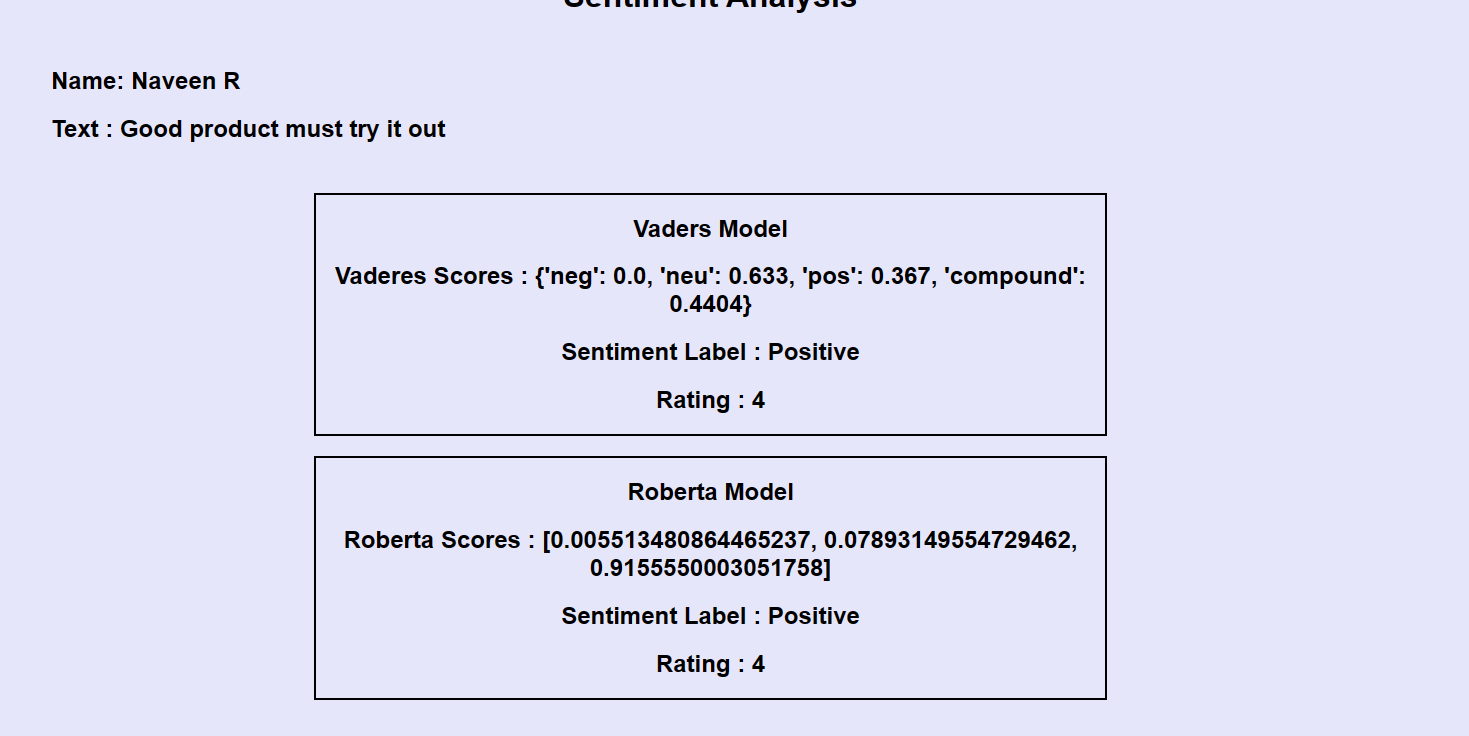
1. Open the deployed web app in a browser. If the deployment is successful the following webpage is rendered.



1. The user is required to input their name and provide a review in the designated fields before clicking on the submit button.



1. After submission, a page is presented exhibiting the outputs of both VADER and Roberta sentiment analysis models. These outputs determine the sentiment label and corresponding rating displayed on the page.



**Conclusion**:

The analysis is based on the customer sentiment for Flipkart product reviews using Varders Model and Roberta Model. The rating distributions showed a positive sentiment with a majority of rating being 5. VADER model provides the negativity, neutrality, positivity and compound score for each review. Roberta Model classifies the reviews into positive, negative or neutral categories with probabilities. By using both models, the strengths of each model was leverage, the VADER provides a quick and interpretable analysis, allowing you to understand the basic sentiment and reasoning behind the classification. Roberta offers a more sophisticated analysis, capturing the subtleties of language and providing probabilities for a more nuanced understanding.

Flask-based web application was developed to generate the real-time sentiment analysis. User can input name and review, and application displays the sentimental analysis from VADER and Roberta models. This sentiment analysis can be a powerful tool for Flipkart to understand customers satisfaction and improve product offerings.