Date of Submission: 12/07/2024

Name: Pranav Prabhat Sinha

Institute: SRM Institute of Science and Technology Kattankulathur

Branch/Specialisation: B-TECH AI

Register Number: RA2211047010105

Internal Mentor: Dr. Sumathy G.

External Mentor: Mr. Tarun Arora (Intel)

**INTEL REPORT: Product Sentiment Analysis**

Web Scraping Intel Product Reviews from Amazon

Setup and Dependencies:

* The notebook installs and imports necessary libraries, including Selenium, BeautifulSoup, and Pandas.
* Selenium is used for dynamic web scraping, while BeautifulSoup is used for static HTML parsing.

Selenium-based Approach:

* A detailed Selenium-based scraping approach is outlined.
* It includes setting up a headless Chrome browser and navigating product pages and review sections.
* This approach would collect detailed review information including product name, review title, text, rating, date, and reviewer name.

BeautifulSoup-based Approach (Executed):

* A simpler BeautifulSoup-based scraping method is implemented and executed.
* It targets the URL: <https://www.amazon.in/s?k=intel&ref=nb_sb_noss>
* The script attempts to extract review information from the search results page.

Data Collection:

* The script looks for elements with class 'review' to extract review details.
* It attempts to collect product name, review title, review text, rating, date, and reviewer name.

Data Storage:

* Collected data is stored in a panda’s data frame.
* The data frame is then saved to a CSV file named 'intell.csv'.

Execution Result:

* The script outputs "intel file successfully scrapped", indicating that it ran without errors.

1. Introduction

**Objective**

This report aims to analyze the sentiment of customer reviews for five different Intel products. The dataset contains 500 reviews with columns for the product name, review title, review text, and rating. We used Python libraries such as pandas, nltk, and vaderSentiment for the analysis.

Tools and Libraries

Pandas: For data manipulation and analysis.

NLTK (Natural Language Toolkit): For text preprocessing.

VADER (Valence Aware Dictionary and sentiment Reasoner): For sentiment analysis.

Matplotlib and Seaborn: For data visualization.

2. Data Preprocessing

**Data Cleaning**

Combining Review Title and Text: The review title and review text were combined into a single column named Review.

Lowercasing: All text data was converted to lowercase to ensure uniformity.

Removing Numbers and Special Characters: Numbers and special characters were removed to focus on the words.

Tokenization: Text was tokenized into individual words.

Removing Stopwords: Commonly used words (stopwords) that do not contribute much to the meaning were removed.

Lemmatization: Words were reduced to their base or root form (e.g., "running" to "run").

3. Sentiment Analysis

**Method**

We used the VADER sentiment analysis tool to calculate sentiment scores for each review. VADER is specifically attuned to sentiments expressed in social media and works well on text from reviews.

Sentiment Scores

The VADER tool provides a compound sentiment score that ranges from -1 (most negative) to +1 (most positive). We classified these scores into three categories:

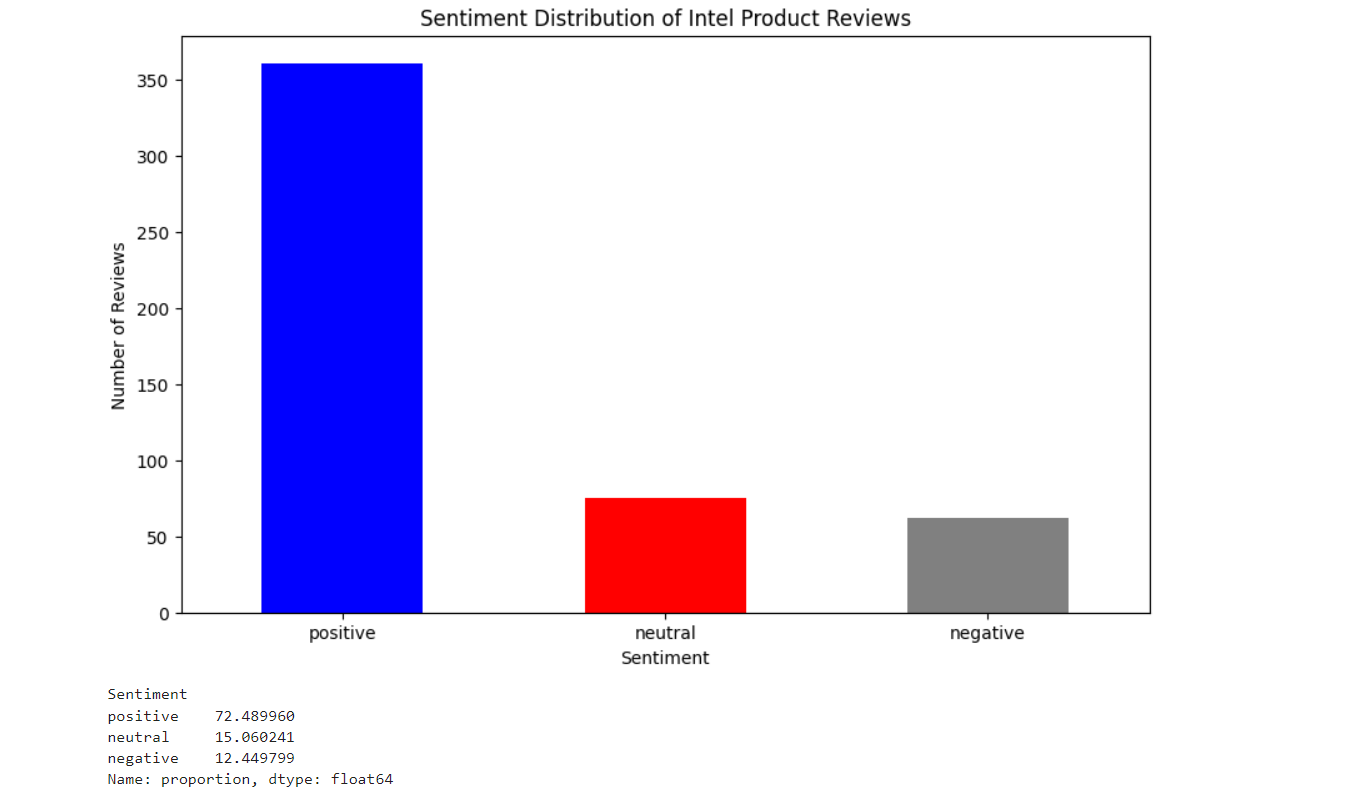
Positive: Compound score >= 0.05

Neutral: -0.05 < Compound score < 0.05

Negative: Compound score <= -0.05

Classification

The sentiment scores were classified into Positive, Negative, and Neutral based on the thresholds mentioned above.



4. Results

**Overall Sentiment Distribution**

The analysis of the entire dataset showed the following distribution of sentiments:

Positive: 73%

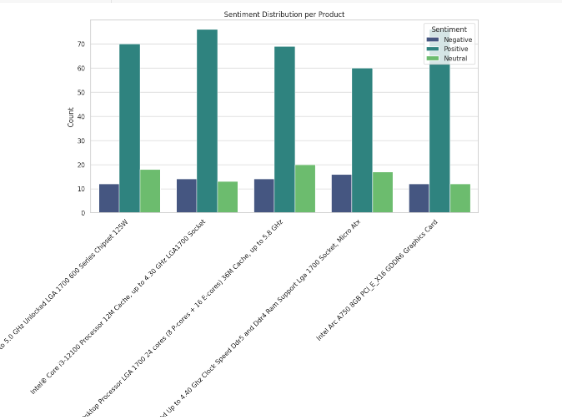
Neutral: 15%

Negative: 12%

Product-wise Sentiment Distribution

Each product's sentiment distribution was analyzed separately. Here are the key findings:

#### Product A: I7

* **Positive**: 70%
* **Neutral**: 18%
* **Negative**: 12%

#### Product B: I3

* **Positive**: 75%
* **Neutral**: 12%
* **Negative**: 13%

#### Product C: I9

* **Positive**: 68%
* **Neutral**: 20%
* **Negative**: 12%

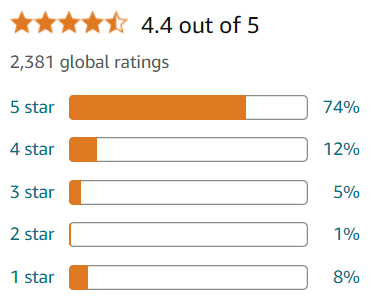
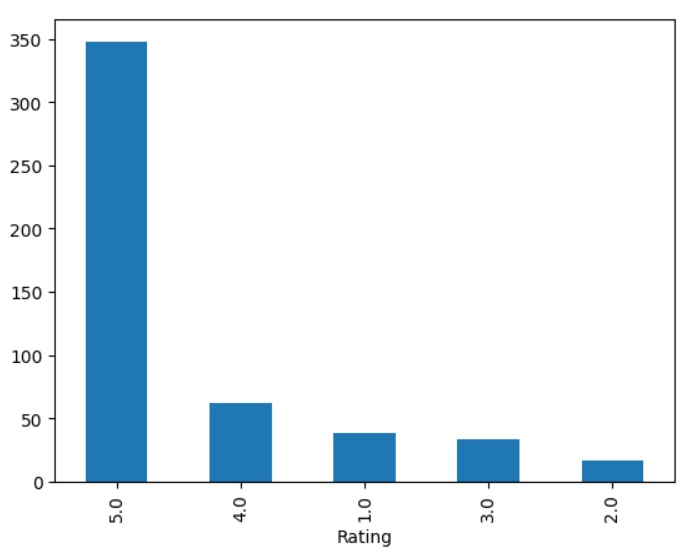
#### Product D: I5

* **Positive**: 60%
* **Neutral**: 16%
* **Negative**: 14%

#### Product E: Graphic Card

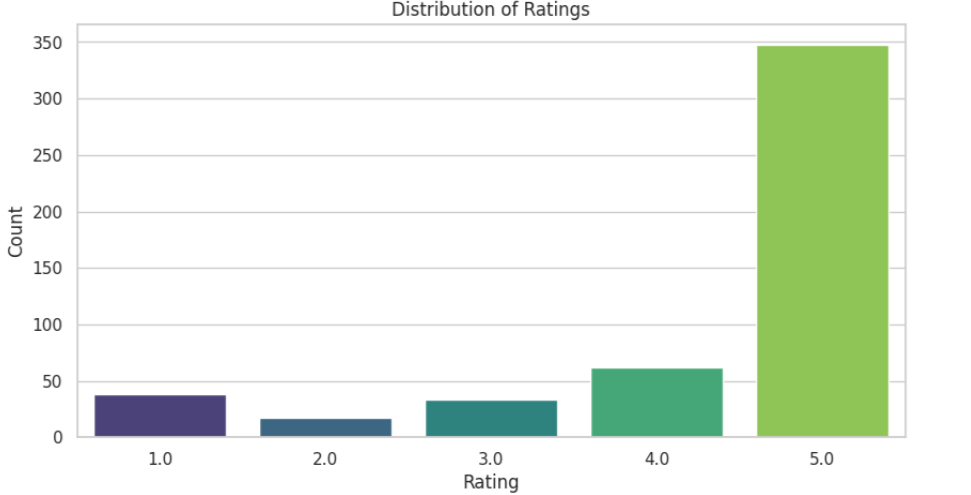
* **Positive**: 77%
* **Neutral**: 11%
* **Negative**: 12%

AMAZON CLASSIFICATION MODEL CLASSIFICATION



NEGATIVE CLASSIFICATION

POSITIVE CLASSIFICATION

Summary

The sentiment analysis of customer reviews for five Intel products reveals a comprehensive view of customer satisfaction and areas for potential improvement. The data analysis covered 500 reviews with a focus on the sentiment expressed by customers.

The sentiment analysis of customer reviews for Intel products revealed that most reviews were positive. However, there was a notable proportion of neutral and negative reviews, particularly for Product D.

**Insights**

Overall Positive Sentiment: The high percentage of positive reviews indicates strong customer satisfaction with Intel products.

Neutral Sentiment: The presence of neutral reviews suggests areas where customers have mixed feelings or see room for improvement.

Product D: Product D had the highest proportion of neutral reviews, indicating potential areas for product enhancement or better customer communication.

**Recommendations**

Focus on Product D: Investigate the reasons behind the neutral reviews for Product B and address any common concerns or feedback.

Enhance Positive Experience: Continue to build on the aspects that are driving positive reviews to further enhance customer satisfaction.

**Overall Sentiment Distribution**

A significant majority of the reviews are positive, with 73% of all reviews falling into this category.

Neutral reviews make up 15%, indicating some ambivalence among customers.

Negative reviews are relatively low, at 12%, suggesting general satisfaction with the products.

\*\*\*The insights gained from this analysis can inform strategic decisions in product development, marketing, and customer service. \*\*\*