# Intel Products Sentiment Analysis from Online Reviews

Intel Unnati Industrial Training Program 2024

**Category A: End Users/Buyers Reviews** 

Name : Sakshi Nirajkumar Rangwala

University: Dhirubhai Ambani Institute of Information and

Communication Technology, Gandhinagar, Gujarat.

Mentor : Manish Kumar Gupta

# **Table of Contents**

Index	Title
1	Abstract
2	Introduction
3	Literature Review
4	Data Collection
5	Data Preprocessing
6	Sentiment Analysis Methodology
7	Implementation
8	Results and Discussion
9	Conclusion
10	References

# **Abstract**

This project presents a comprehensive sentiment analysis of Amazon reviews for Intel processors. The primary objectives of this project are to extract, clean and analyze customer feedback to gain insights into consumer sentiment and provide actionable suggestions for product improvement.

The methodology involves extracting data from Amazon using web scraping techniques and cleaning the data through a series of preprocessing steps, including removing duplicates, handling missing values and standardizing text. Sentiment analysis is conducted using natural language processing (NLP) techniques to categorize the reviews as positive, negative or neutral. The analysis reveals prevalent trends and common issues faced by customers.

Based on the findings, I offer specific recommendations to enhance customer satisfaction and product performance. Additionally, I have developed APIs to facilitate the data cleaning and analysis processes, ensuring the methodology can be easily replicated and applied to other datasets.

In conclusion, this project demonstrates the value of sentiment analysis in understanding customer feedback and driving product improvements. The developed APIs further enhance the utility of this analysis, providing a scalable solution for ongoing sentiment evaluation.

# **Introduction**

#### **Project Background:**

User reviews have become an essential component in evaluating the performance and satisfaction associated with consumer products. In the realm of technology, particularly for products like Intel processors, user reviews provide invaluable insights into real-world performance, usability, and potential issues that might not be evident through traditional testing methods. These reviews, sourced from platforms like Amazon, reflect genuine user experiences and opinions, offering a rich dataset for sentiment analysis. Understanding and analyzing these sentiments can help Intel and other stakeholders identify strengths and weaknesses, thereby guiding product development and customer support strategies.

#### **Objective:**

The objective of this project is to discern overall consumer sentiment, identify common themes in reviews and provide actionable recommendations to improve product quality and customer satisfaction.

#### Scope:

The scope of this project includes the extraction of reviews specifically from Amazon's platform, focusing on Intel processor products. The analysis encompasses reviews posted within the past four years to ensure the data reflects current consumer opinions and trends. By limiting the timeframe, we aim to capture the most relevant and up-to-date feedback, providing accurate and timely insights for Intel's product development and marketing strategies.

### **Literature Review**

#### **Related Work:**

Sentiment analysis has been extensively studied in the context of product reviews, with numerous research efforts focusing on extracting insights from customer feedback to inform product development and marketing strategies. Previous studies have demonstrated the utility of sentiment analysis in various domains, including electronics, fashion and consumer goods. In the tech industry, sentiment analysis of product reviews has been particularly valuable, given the rapid pace of innovation and the critical role of customer feedback in shaping product iterations.

Several notable studies have focused on sentiment analysis of tech products, including processors. For instance, research by Hu and Liu (2004) pioneered the use of sentiment analysis for mining opinions from online product reviews, setting the groundwork for subsequent studies. Their approach involved identifying opinion words and determining their polarity to classify sentiments as positive, negative or neutral. More recent work by Zhang et al. (2018) applied deep learning techniques to enhance sentiment classification accuracy in tech product reviews, highlighting the evolution of methodologies in this field.

In the context of Intel processors, sentiment analysis has been employed to understand user satisfaction and identify common issues such as performance bottlenecks, compatibility concerns and thermal management. These insights have been instrumental in guiding product enhancements and addressing customer pain points effectively.

# **Sentiment Analysis Techniques:**

Sentiment analysis encompasses a variety of techniques and tools, each with its own strengths and applications. The most common approaches include:

- Lexicon-Based Methods: These techniques rely on predefined dictionaries of sentiment-laden words (lexicons) to determine the sentiment polarity of a text. Examples include SentiWordNet and the AFINN lexicon. While lexicon-based methods are straightforward and interpretable, they may struggle with context-specific nuances and slang.
- 2) **Machine Learning-Based Methods**: These involve training classifiers, such as Support Vector Machines (SVM), Naive Bayes or Random

- Forests, on labelled datasets to predict sentiment. Machine learning models can capture complex patterns in data but require substantial labelled data for training and may not generalize well to unseen data.
- 3) **Deep Learning-Based Methods**: Recent advancements in deep learning have significantly improved sentiment analysis accuracy. Techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformers (e.g., BERT) have demonstrated superior performance by capturing intricate language patterns and contextual information. These models, however, demand considerable computational resources and large datasets for training.
- 4) **Hybrid Methods**: Combining lexicon-based and machine learning approaches can leverage the strengths of both techniques. For instance, lexicons can provide initial sentiment scores, which are then refined using machine learning models.

In this project, we employ the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis technique, a lexicon and rule-based model specifically attuned to sentiments expressed in social media and product reviews. VADER is particularly effective for analyzing text from online reviews due to its ability to handle both conventional and informal language. Additionally, we integrate APIs for data cleaning and analysis, ensuring scalability and reproducibility of the methodology, which enables ongoing monitoring of customer sentiment.

By reviewing these related works and techniques, I establish a solid foundation for my sentiment analysis project, ensuring that my approach leverages the latest advancements and best practices in the field. Utilizing VADER for sentiment analysis, combined with robust data processing through APIs, provides an efficient and accurate means to derive actionable insights from Intel processor reviews.

# **Data Collection**

In this project, I collected over 3000 reviews of Intel processors from Amazon to perform a comprehensive sentiment analysis. The data collection process was meticulously designed to ensure the accuracy and relevance of the extracted information.

To achieve this, I utilized two powerful Python libraries: BeautifulSoup and Requests.

#### **Tools and Techniques:**

**BeautifulSoup**: This Python library was used to parse the HTML content of the Amazon review pages. BeautifulSoup enabled us to navigate the HTML structure and extract specific elements containing review data.

**Requests**: To fetch the HTML content of the Amazon review pages, the requests library was used. This allowed us to make HTTP requests and retrieve the page content, which was then parsed using BeautifulSoup.

#### **Steps Involved in Data Collection:**

#### 1) Review URL Collection:

- a) I started by identifying the URLs of the product pages containing Intel processor reviews.
- b) All review URLs were stored in a single text file for sequential processing.

#### 2) Data Extraction:

- a) Using the requests library, I loaded each URL from the text file.
- b) BeautifulSoup was then employed to parse the HTML content of each loaded page, extracting relevant information.
- c) The extracted data included the following columns:
  - i) **Customer Name**: The name or username of the reviewer.
  - ii) **Rating**: The star rating given by the customer (ranging from 1 to 5 stars).
  - iii) Review Title: A brief title summarizing the review.
  - iv) **Review Text**: The detailed text of the review.
  - v) **Review Date**: The date when the review was posted.
  - vi) **Product Name**: The name of the Intel processor being reviewed.
- d) If the 'Review Text' column is in the other language than English, I translated the text into English for better analysis.

#### 3) Time Frame:

a) To ensure the data's relevance and reflect recent consumer opinions, I extracted reviews posted within the last four years.

#### 4) Data Storage:

- a) After extracting the data from each review page, I compiled all the information into a CSV file.
- b) This structured format facilitated easy manipulation, analysis and sharing of the data.

The combination of BeautifulSoup and Requests proved effective for handling the dynamic nature of Amazon's review pages, allowing us to systematically collect a substantial dataset for sentiment analysis. This dataset serves as the foundation for our subsequent data cleaning, analysis and recommendation processes, providing valuable insights into consumer sentiments regarding Intel processors.

# **Data Preprocessing**

Data preprocessing is a critical step in preparing the extracted dataset for analysis. It involves cleaning and transforming raw data to ensure it is in the optimal format for sentiment analysis. In this project, the preprocessing steps included handling missing values, extracting relevant information from existing columns, removing duplicates and categorizing product names. Here's a detailed account of the preprocessing steps undertaken:

#### 1) Handling Missing Values:

- a) Dropping Column 'review\_title' with Most Null Values: The column 'review\_title' with the most null values was identified and removed from the dataset to reduce noise and improve data quality.
- b) **Deleting Rows with Null Values in 'review\_text' column**: Since the 'review\_text' column is crucial for sentiment analysis, any rows with null values in this column were deleted. This ensured that only complete reviews were included in the analysis.

#### 2) Extracting and Transforming Date Information:

- a) Adding 'date' and 'country' Columns: From the 'review\_date' column, two new columns were derived. The 'date' column extracted the specific date of the review, while the 'country' column identified the country from which the review was posted.
- b) **Deleting 'review\_date' Column**: After extracting the necessary information, the original 'review\_date' column was deleted to avoid redundancy and streamline the dataset.

#### 3) Removing Duplicate Rows:

a) Duplicate rows were identified and removed to ensure that each review was unique. This step was essential to avoid bias and over-representation of certain reviews in the analysis.

#### 4) Classifying Product Names into Categories:

a) The product names were classified into predefined categories to facilitate a more organized and systematic analysis. This classification helped in aggregating reviews for similar products and identifying trends and patterns within each category.

# **Sentiment Analysis Methodology**

For this project, I adopted a rule-based approach to sentiment analysis using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool. VADER is a lexicon and rule-based sentiment analysis tool that is specifically designed to analyze sentiments expressed in social media texts. Developed by researchers Hutto and Gilbert, VADER is particularly effective in handling the nuances of sentiment analysis in informal texts such as tweets, online reviews, and other short texts.

#### **Functionality of VADER:**

- Lexicon-based Analysis: VADER employs a pre-built lexicon (dictionary) containing sentiment scores for words and phrases. Each word in the lexicon is assigned a polarity score (positive, negative, or neutral) and an intensity score which determines the strength of the sentiment.
- 2) Rule-based Analysis: In addition to the lexicon, VADER uses a set of grammatical and syntactical rules to interpret the sentiment expressed in a text. These rules help VADER understand the context and structure of sentences to determine sentiment more accurately.
- 3) **Polarity Detection**: VADER is capable of detecting the polarity of sentiments expressed in text, categorizing them as positive, negative or neutral. It provides a sentiment intensity score for each category, indicating the strength of the sentiment expressed.
- 4) Emoticon and Emoji Handling: VADER is designed to handle emoticons and emojis commonly used in social media texts. It assigns sentiment scores to emoticons and emojis based on their context and usage in the text.
- 5) **Handling of Intensifiers and Negations**: VADER considers the effect of intensifying words (e.g., "very", "extremely") and negations (e.g., "not", "never") on sentiment analysis. It adjusts the sentiment scores accordingly to account for the degree of intensity and negation in the text.

- 6) **Sentence-level Analysis**: VADER analyzes sentiment at the sentence level, allowing it to capture the sentiment expressed in individual sentences within a text. This fine-grained analysis helps in understanding the overall sentiment of the text more accurately.
- 7) Compound Sentiment Score: VADER calculates a compound sentiment score for the entire text, which represents the aggregated sentiment intensity. The compound score provides a single metric for overall sentiment, considering both positive and negative sentiments expressed in the text.

Overall, VADER offers a robust and efficient solution for sentiment analysis in social media texts, enabling researchers and analysts to gain insights into the sentiment expressed by users in online platforms. Its combination of lexicon-based and rule-based approaches, along with its handling of emoticons, intensifiers and negations, makes it well-suited for analyzing sentiments in informal text data.

#### VADER uses **Polarity and Intensity Scores**:

- Polarity Score: Indicates sentiment orientation, ranging from -1 (negative) to +1 (positive).
- **Intensity Score**: Indicates the strength of the sentiment, with higher values reflecting stronger sentiment.

# How VADER identifies text as positive, negative or neutral?

VADER (Valence Aware Dictionary and sEntiment Reasoner) identifies text as positive, negative, or neutral based on the sentiment scores assigned to individual words and phrases in the text. Here's how VADER determines the sentiment of a piece of text:

- 1) **Lexicon Lookup**: VADER uses a pre-built lexicon (dictionary) containing thousands of words and phrases, each with an associated polarity score. These scores indicate whether a word or phrase is positive, negative, or neutral, as well as the intensity of the sentiment.
- 2) **Sentiment Intensity Scores**: When analyzing a piece of text, VADER breaks it down into individual words and phrases and looks up their sentiment scores in the lexicon. It calculates the sentiment intensity

- score for each word or phrase based on its polarity score and the context in which it appears in the text.
- 3) Aggregation of Scores: VADER aggregates the sentiment intensity scores of all words and phrases in the text to calculate overall sentiment scores for positive, negative, and neutral sentiments. It considers both the polarity and intensity of the sentiments expressed by individual words and phrases.
- 4) Thresholds: VADER applies predefined thresholds to the aggregated sentiment scores to classify the overall sentiment of the text. If the positive sentiment score exceeds a certain threshold, the text is classified as positive. Similarly, if the negative sentiment score exceeds a threshold, the text is classified as negative. If neither positive nor negative sentiment scores meet their respective thresholds, the text is classified as neutral.
- 5) Handling of Emoticons and Emojis: VADER also considers emoticons and emojis present in the text. It assigns sentiment scores to these symbols based on their typical connotations and usage, which contributes to the overall sentiment analysis.
- 6) Consideration of Intensifiers and Negations: VADER takes into account the effect of intensifying words (e.g., "very", "extremely") and negations (e.g., "not", "never") on sentiment analysis. It adjusts the sentiment scores accordingly to account for the degree of intensity and negation in the text.

By combining lexicon-based sentiment analysis with rule-based processing and considering various linguistic features, VADER can effectively identify the sentiment expressed in a piece of text and classify it as positive, negative or neutral.

# <u>Implementation</u>

In the implementation phase of our sentiment analysis project, several key steps were undertaken to process the data, analyze sentiments and derive actionable insights. This section outlines the specific methodologies and technologies employed to achieve our objectives.

#### **Tools and Libraries:**

The implementation of sentiment analysis using VADER was carried out using the following tools and libraries:

- **Python**: The primary programming language for data processing and analysis.
- NLTK (Natural Language Toolkit): Used for tokenization and accessing VADER.
- pandas: For data manipulation and cleaning.
- matplotlib and seaborn: For data visualization.
- FastAPI: For creating APIs to facilitate data cleaning and analysis.
- **Redis**: Used as a server for efficient data storage and retrieval.

# **Sentiment Analysis:**

- **Sentiment Column Addition**: Using the VADER sentiment analysis tool, a new column was added to the dataset to classify each review as positive, negative or neutral based on the 'review\_text' column. This provided a straightforward categorization of customer sentiments.
- Aspect Extraction: Aspects were extracted from the 'review\_text'
  column to identify specific features or attributes of the Intel processors
  that customers mentioned. This involved identifying nouns and noun
  phrases that represent aspects such as performance, price, durability,
  etc.
- Aspect Sentiment: Alongside aspect extraction, the sentiment associated with each aspect was also determined. This allowed us to understand not only what features customers discussed but also how they felt about these features.

#### **Key Feature Analysis:**

- Key Feature Extraction: The most frequently mentioned aspects were identified and labelled as key features. This helped in focusing on the most relevant aspects that customers care about.
- Insights on Strengths and Areas for Improvement: By analyzing the sentiment associated with the key features, insights were derived regarding the strengths of the Intel processors and the areas that need improvement. Positive sentiments highlighted strengths such as high performance and reliability, while negative sentiments pointed to issues like high price or heat generation.

#### **Unique Features:**

#### **API Development:**

To streamline the data cleaning and analysis processes, APIs were developed using FastAPI and Redis. This approach allowed for efficient and scalable data processing, facilitating ongoing sentiment analysis and aspect extraction as new reviews are collected. Additionally, a user interface (UI) page was created to provide easy access to these APIs.

- **FastAPI**: Used to create RESTful APIs for various functions, including data cleaning and sentiment analysis. These APIs enable automated processing of new review data, making the system scalable and easy to integrate with other applications.
- Redis: Implemented as a caching server to store intermediate results and frequently accessed data, significantly improving the performance and responsiveness of the APIs.
- User Interface (UI) Page: A UI was developed to allow users to easily interact with the APIs. This page provides a user-friendly interface for accessing the data cleaning and sentiment analysis functionalities, ensuring that the system is accessible to users without technical expertise.

# Implementation Workflow:

1) **Data Cleaning APIs**: API endpoints were created to handle data cleaning tasks, such as removing duplicates, handling missing values, and extracting relevant columns. This ensures that the data fed into the analysis pipeline is clean and consistent.

- 2) **Sentiment Analysis APIs**: Another endpoints were developed to perform sentiment analysis on the 'review\_text' column, adding the sentiment classification to the dataset.
- 3) **Aspect Extraction API**: This endpoint extracts aspects and their associated sentiments from the reviews, storing the results in a structured format for further analysis.
- 4) **Insights and Reporting**: The cleaned and analyzed data was used to generate visualizations, providing insights into customer sentiments and key features. These visualizations were essential for understanding customer feedback and identifying areas for product improvement.

# **Results and Discussions**

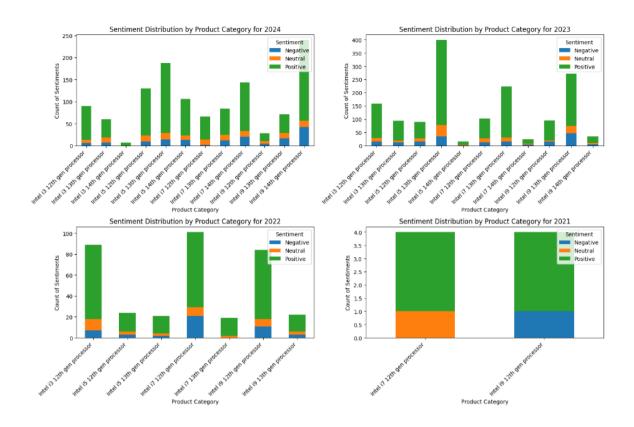
# Category wise sentiment counts

#### Negative Neutral Positive category Intel i3 12th gen processor Intel i3 13th gen processor Intel i3 14th gen processor Intel i5 12th gen processor Intel i5 13th gen processor Intel i5 14th gen processor Intel i7 12th gen processor Intel i7 13th gen processor Intel i7 14th gen processor Intel i9 12th gen processor Intel i9 13th gen processor Intel i9 14th gen processor

# Category wise rating statistics

	mean	median	std
category			
Intel i3 12th gen processor	4.68	5.00	0.82
Intel i3 13th gen processor	4.57	5.00	1.01
Intel i3 14th gen processor	5.00	5.00	0.00
Intel i5 12th gen processor	4.30	5.00	1.30
Intel i5 13th gen processor	4.76	5.00	0.72
Intel i5 14th gen processor	4.59	5.00	0.93
Intel i7 12th gen processor	4.56	5.00	1.08
Intel i7 13th gen processor	4.69	5.00	0.81
Intel i7 14th gen processor	4.41	5.00	1.16
Intel i9 12th gen processor	4.46	5.00	1.18
Intel i9 13th gen processor	4.27	5.00	1.37
Intel i9 14th gen processor	4.35	5.00	1.24

# Product category wise sentiment counts per year :



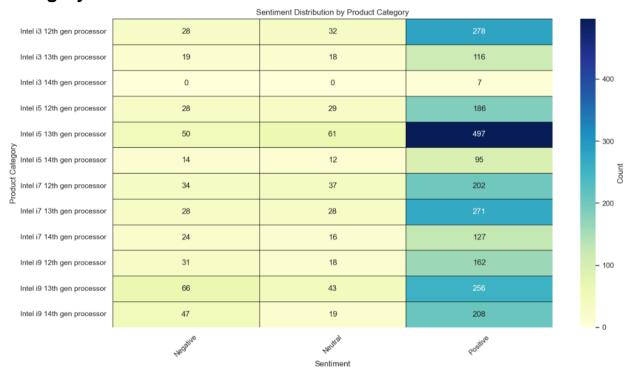
# **Category wise Customer Satisfaction Percentage and Negative Reviews Percentage**

Category	Customer Satisfaction Percentage	Negative Reviews Percentage
Intel i3 12th gen processor	82.25%	8.28%
Intel i3 13th gen processor	75.82%	12.42%
Intel i3 14th gen processor	100.00%	0.00%
Intel i5 12th gen processor	76.54%	11.52%
Intel i5 13th gen processor	81.74%	8.22%
Intel i5 14th gen processor	78.51%	11.57%
Intel i7 12th gen processor	73.99%	12.45%
Intel i7 13th gen processor	82.87%	8.56%
Intel i7 14th gen processor	76.05%	14.37%
Intel i9 12th gen processor	76.78%	14.69%
Intel i9 13th gen processor	70.14%	18.08%
Intel i9 14th gen processor	75.91%	17.15%

#### Insights:

- The **Intel i3 14th gen processor** stands out with the highest customer satisfaction (100.00%) and the lowest negative reviews percentage (0.00%).
- The Intel i9 13th gen processor has the lowest customer satisfaction (70.14%) and the highest negative reviews percentage (18.08%).

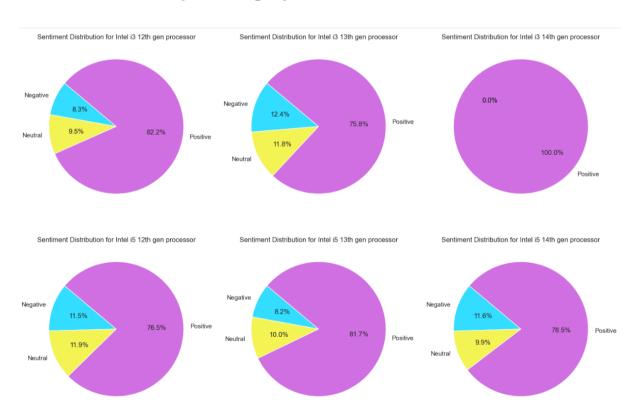
# **Category wise sentiment distribution**

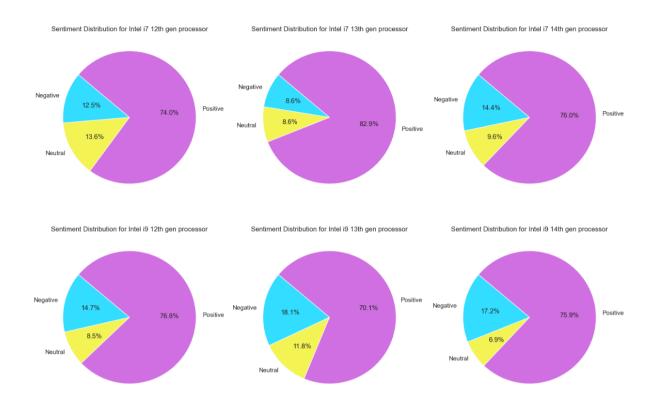


#### Year wise sentiment distribution by product category



# Sentiment counts per category

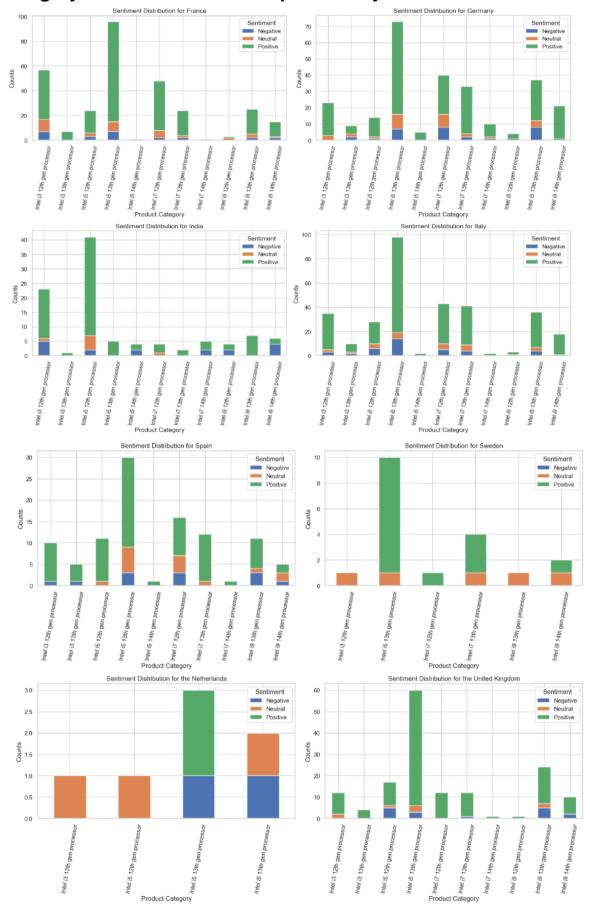




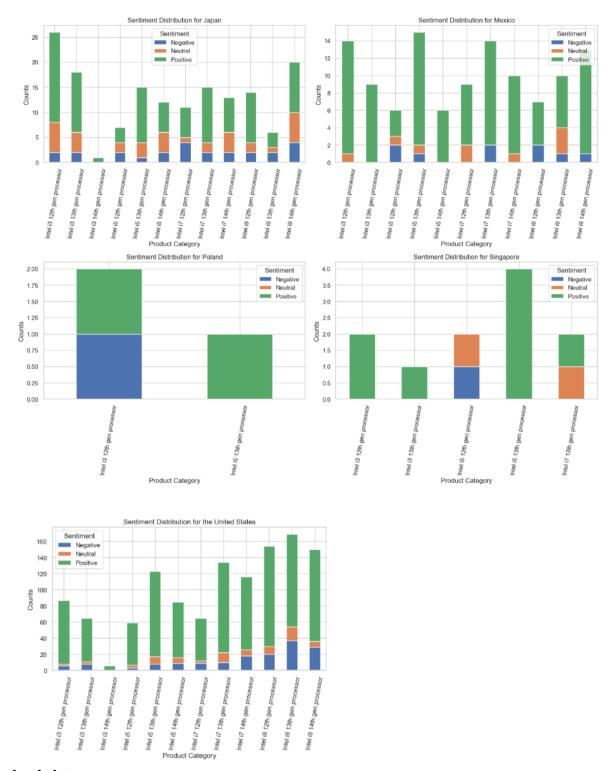
# Insights:

- The Intel i3 14th gen processor has all positive reviews.
- The Intel i9 13th gen processor has the highest negative reviews. Also Intel i9 processors have more negative reviews compared to other processors.

# Category wise sentiment count per country



Page 21 of 37



# Insights:

• People of United States are more active in giving reviews whereas for the Poland, there is exactly the opposite case.

# Improvements needed for the features :

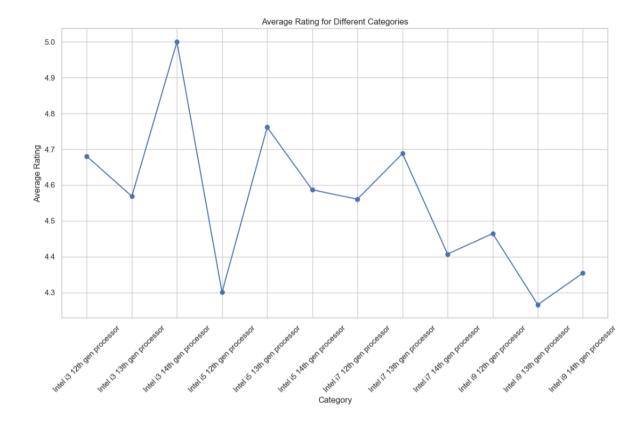
	Count	Percentage
Key Feature		
сри	119	32.425068
processor	71	19.346049
performance	27	7.356948
k	23	6.267030

# Keyfeature strengths:

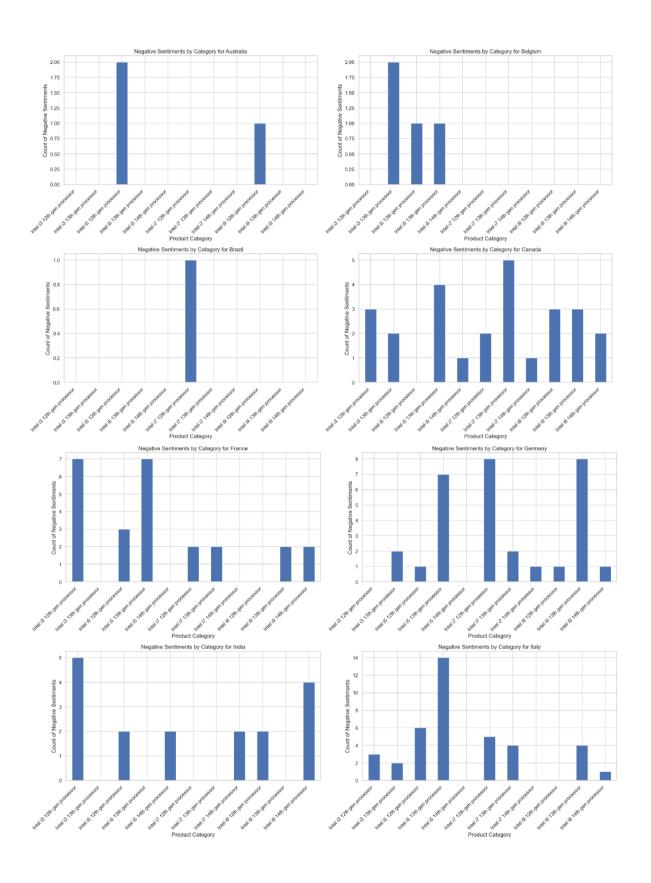
# Frequency Percentage (%)

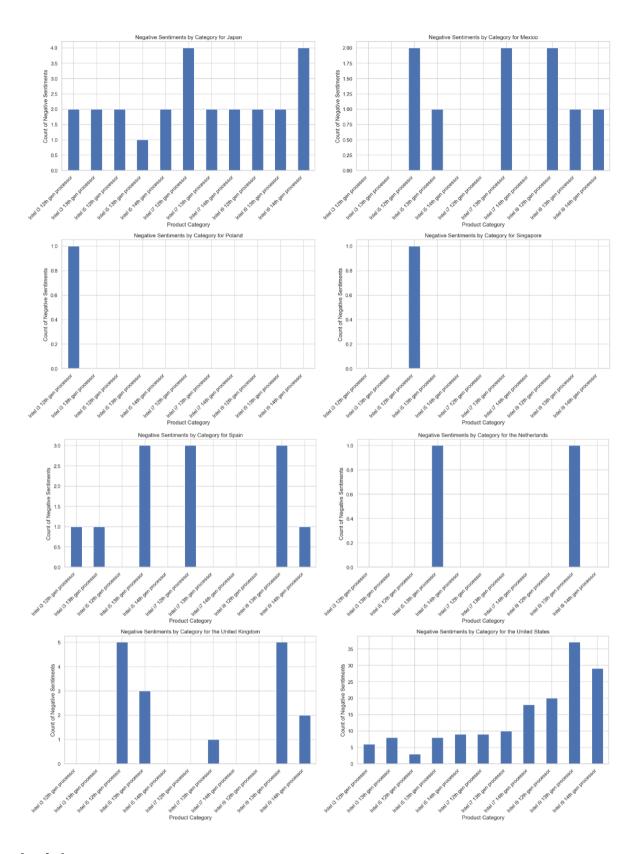
Key Feature		
cpu	693	30.13%
processor	530	23.04%
performance	171	7.43%
good	117	5.09%
k	79	3.43%
intel	65	2.83%
gaming	62	2.70%
price	59	2.57%
product	36	1.57%
games	31	1.35%
cooler	29	1.26%
i	25	1.09%
time	20	0.87%
рс	18	0.78%
great	18	0.78%
excellent	13	0.57%
power	11	0.48%
new	11	0.48%

# Average rating for different categories



# **Category wise negative sentiments per country**

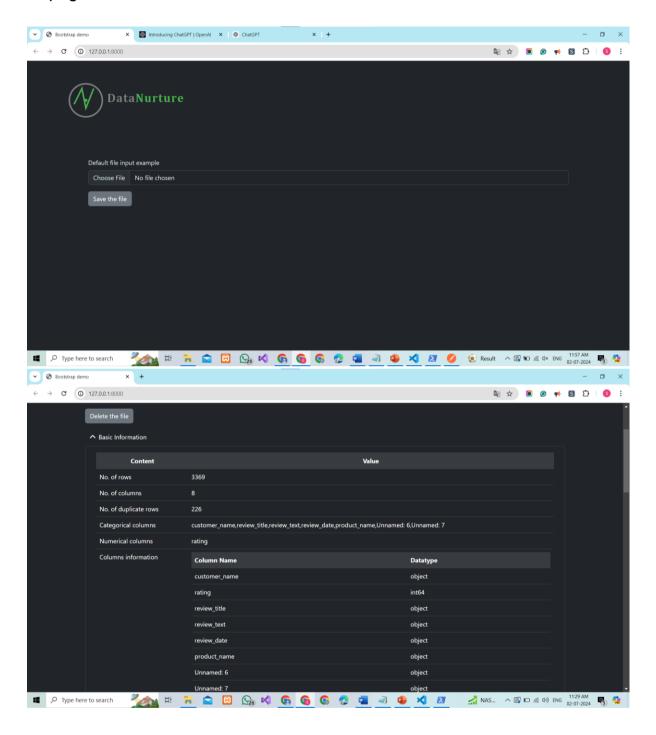


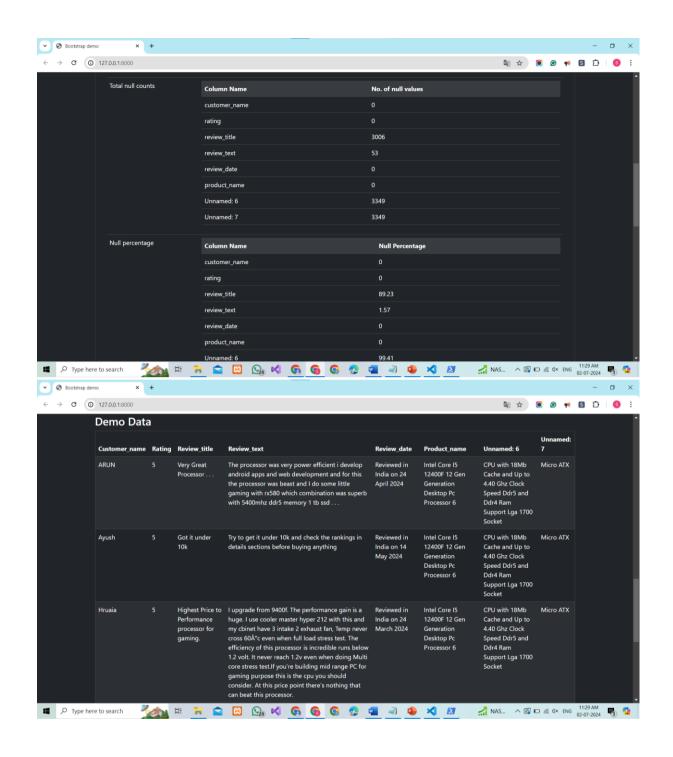


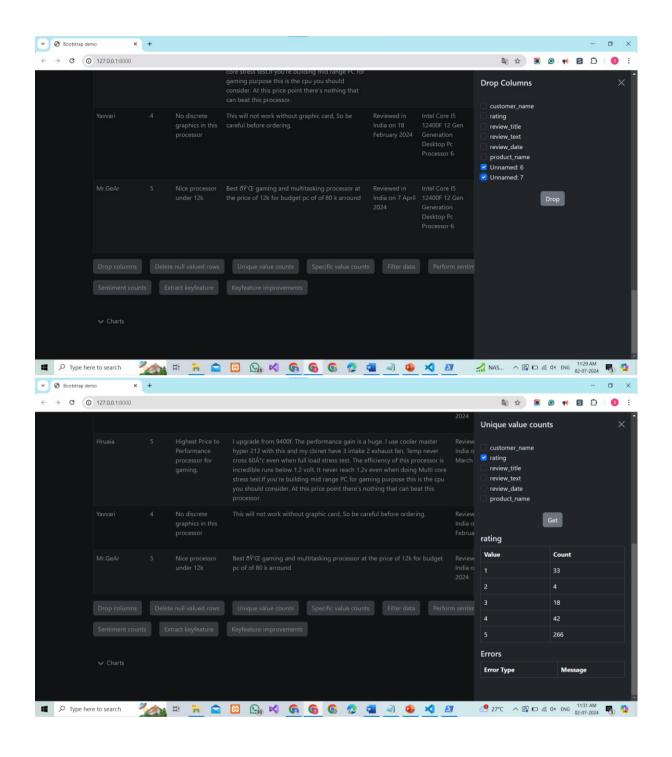
# Insights:

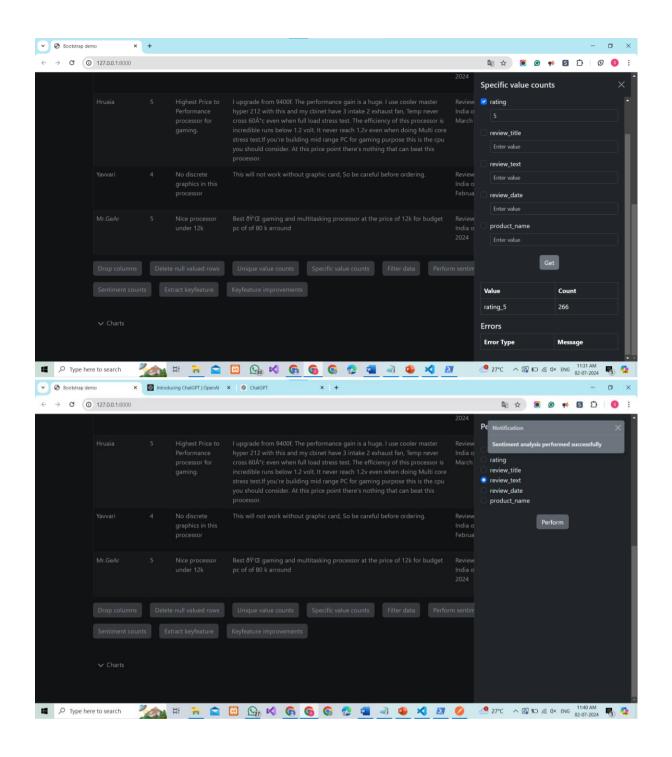
- The United States have highest negative reviews.
- Brazil, Poland and Singapore have only one negative review.

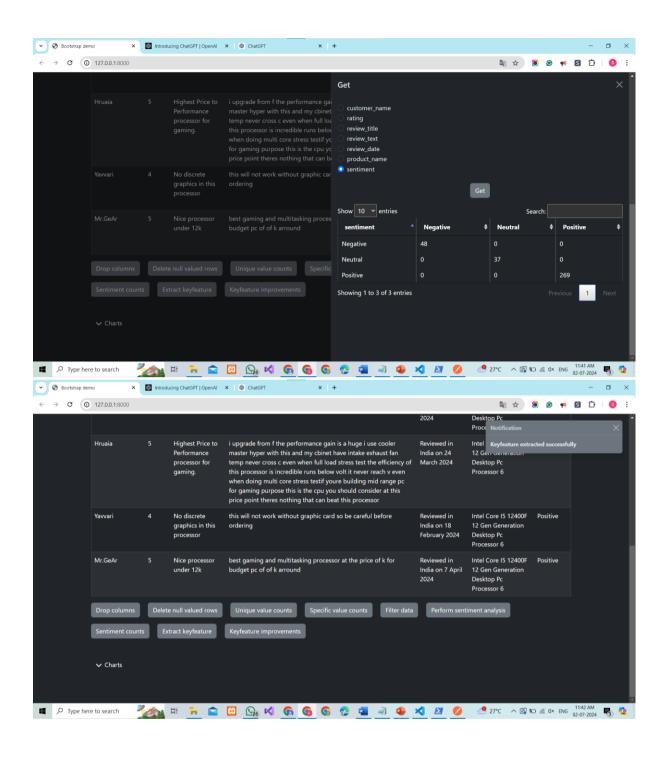
#### UI page to work with APIs

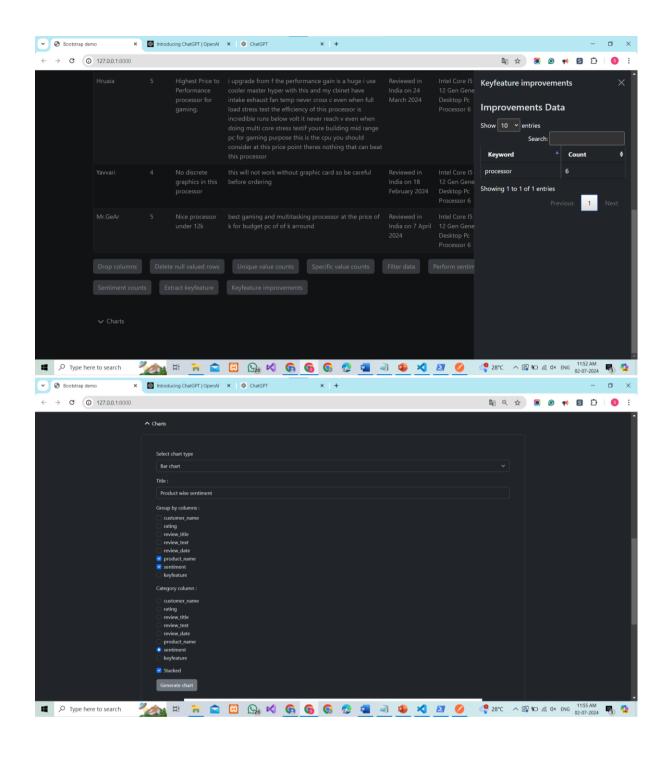


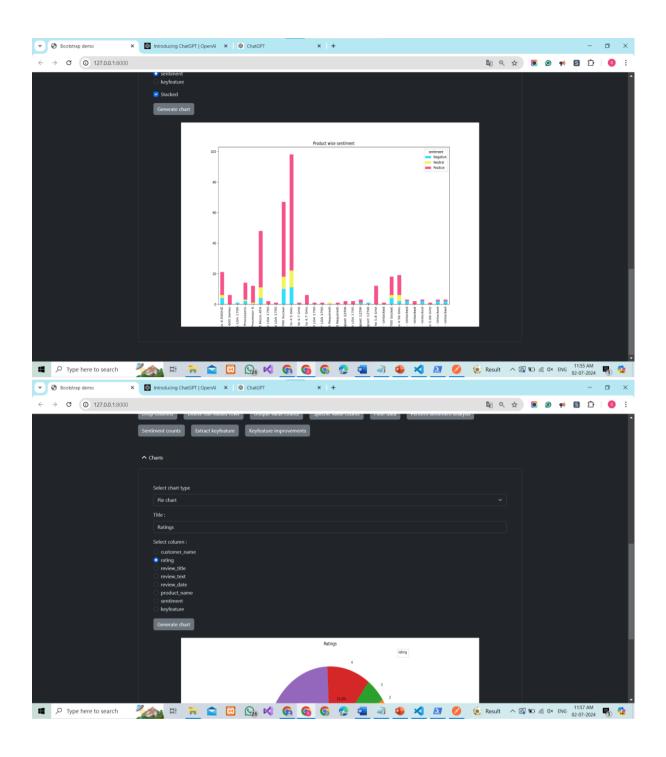


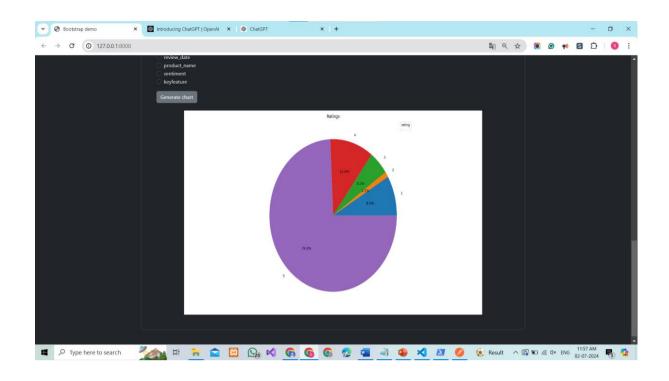












# **Conclusion**

#### Summary:

This project analyzed customer sentiments extracted from Amazon reviews focusing on Intel processors. Key findings reveal that the terms CPU, processor and performance consistently emerge as areas needing improvement, while CPU, processor, performance, price and power are perceived strengths. Notably, the Intel i9 processors exhibit significant negative sentiments, with the 13th generation showing the lowest customer satisfaction at 70.14% and the highest negative reviews percentage at 18.08%. In contrast, the Intel i3 14th gen processor stands out with 100.00% customer satisfaction and 0.00% negative reviews, suggesting robust customer approval. The standard deviation observed in customer satisfaction for the Intel i9 13th gen processor underscores the variability in customer experiences and perceptions.

**Implications**: These findings suggest that while Intel processors generally excel in core features like CPU, processor and performance, there are significant disparities in customer satisfaction across different models. Addressing the issues highlighted by negative sentiments towards the Intel i9 processors could potentially improve overall customer satisfaction and mitigate negative feedback.

**Recommendations**: To enhance customer satisfaction and address negative sentiments, Intel could consider focusing on improving the specific areas identified as weaknesses, such as optimizing performance, refining pricing strategies and ensuring consistent power efficiency across their product lines. Future studies could delve deeper into understanding the factors contributing to variability in customer satisfaction and explore innovative approaches to enhance product performance and customer experience.

### References

- 1) Amazon Data extracted from Amazon reviews. Retrieved from Amazon website <a href="https://www.amazon.in/">https://www.amazon.in/</a>
- 2) freecodeCamp *Python Data Analysis*. YouTube. Retrieved from <a href="https://www.youtube.com/playlist?list=PLWKjhJtqVAblvI1i46ScbKV2jH1">https://www.youtube.com/playlist?list=PLWKjhJtqVAblvI1i46ScbKV2jH1</a> gdL7VQ
- 3) Code with Harry Learn FastAPI in 60 Minutes. YouTube. Retrieved from FastAPI Tutorial in Hindi | Lets build a Notes app using FastAPI, MongoDB & Bootstrap
- 4) OpenAI. (2024). ChatGPT for error solving. Retrieved from <a href="https://chat.openai.com/">https://chat.openai.com/</a>