



SRM

INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

PS-11 Intel Products Sentiment Analysis



Done by: Raagashree M
(RA2211030010040)

College Mentor: Dr. M. Safa, Assistant Professor, Department of NWC
Intel Mentor: Mr. Debdyut Hazra, Intel Team

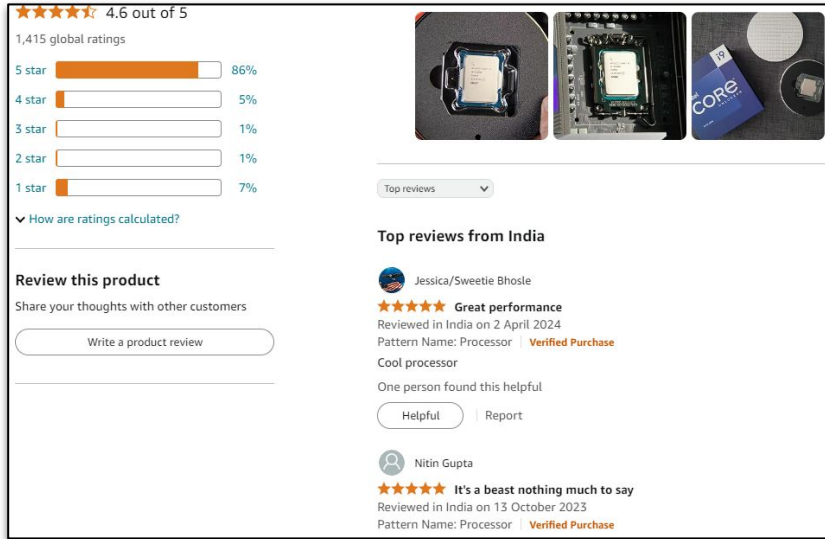
Problem Statement

User reviews give detailed and insightful feedback on the performance, reliability and overall user satisfaction of a product and any issues that are not always mentioned in professional reviews. With the help of these user reviews, it is possible to analyse the user sentiments of specific products and evaluate their performance in the consumer market. Further, it helps analyse the performance analytics of that product and the future improvements to be made, for a good response from the consumer market.

APPROACH



Data collection



To perform sentiment analysis on user sentiments on Intel products (mobile and desktop processors), data was collected as follows:

- **Source:**
The user reviews were sourced from the ecommerce platform, Amazon
- **Tool:**
The user reviews were scraped using an online reviews scraper API provided by Apify.com
- **Dataset:**
Comprised of 2000 reviews categorised by the Intel generations. It includes key features like reviewId, rating score, product Asin and review description

Id														
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Id	reviewId	country	ratingScore	productId	reviewTitle	reviewDescription								
1	R2C3IAC0	Spain	2	B09MDGK	Caja	La caja estaba rota por las esquinas y en malas condiciones, entiendo que sea de la empresa d								
2	R3QN2OX	Sweden	2	B09MDGK	kommer s	stod när jag beställde att den skulle komma 5 januari sen dagen efter kollar jag och ser att den								
3	R121NP83	India	4	B09MDFH	No discre	This will not work without graphic card, So be careful before ordering.								
4	RQKM3JG	India	4	B09MDFH	Nice									
5	R2RY66BJ	India	4	B09MDFH	Keep in m	This one has no graphics processing unit and hence if you use this without graphics card, the b								
6	R3N303EU	India	4	B09MDFH	Good proc	After installing it, Windows 11 is giving the following description of the processor:12th Gen In								
7	R1YIG2BG	India	4	B09MDFH	Great proc	The cpu is overall great for everything, it'll give you flaw less performance in mediocre level v								
8	R3UOVR2	France	4	B09MDFH	Rapport q	Pas cher et efficace pour gaming ,multimédia.								
9	R38CAS2V	France	4	B09MDFH	i5-12400F	Le i5-12400F est un bon processeur avec un prix relativement bas pour ces capacités.Des jeux								
10	R3QCD6W	United Kir	4	B09MDFH	Ideal as ar	Would be useful to specify which boards it will work with as plug and play no longer exist.								
11	RELWSUT	Germany	4	B09MDFH	Absolut B	Ich bin absolut begeistert von der CPUFür den Preis gibt es in Sachen gaming nichts zu mecker								
12	R29GD0W	Italy	4	B09MDFH	ottimo	Buon processore abbinato con una 3070								
13	R1DAGTN	India	4	B09NMPD	Good CPU	Getting optimal speed without any overheating issue								
14	R3LU2TVE	United Sta	4	B09NMPD	Does not	The CPU throttles for no reason during gameplay causing multiple crashes.I would not recomn								
15	R96Q39BC	United Sta	4	B09NMPD	Don't exp	The default cooler can't maintain a reasonable temp, and I found it nearly impossible to instal								
16	R27FZR84	Japan	4	B09NMPD	期待通り	特に問題なく稼働中です。								
17	R3OJL3H5	Canada	4	B09NMPD	Pretty dec	This is pretty decent speed it is as fast or faster than my i7-11700 that runs windows server. Fir								
18	R2X5HXGT	United Sta	4	B09NMPD	Intel Core	Uno de los mejores procesadores para juegos en el 2023 en calidad precio.								
19	R1Y8XVNS	Canada	4	B09NMPD	Box was d	Building a gaming pc, everything is just fine except for the packaging.								
20	R3AITECB	United Sta	4	B09NMPD	Working e	Assisted in solving the problems I had.								

Raw Data collected

Data preprocessing

Data cleaning:

- Removal of duplicate entries based on Review Id to avoid redundancy.
- Handling of missing values.
- Eliminating punctuation marks.
- Translating reviews written in languages other than English to maintain consistency for sentiment analysis and feature extraction.



Text processing:

- Case Normalization
- Tokenization
- **Stop Words Removal:** Removed common words that do not contribute significant meaning to the text.
- Non-Alphabetic Tokens Removal
- **Lemmatization:** Converted words to their base or root form to reduce word variations and enhance text consistency.
-



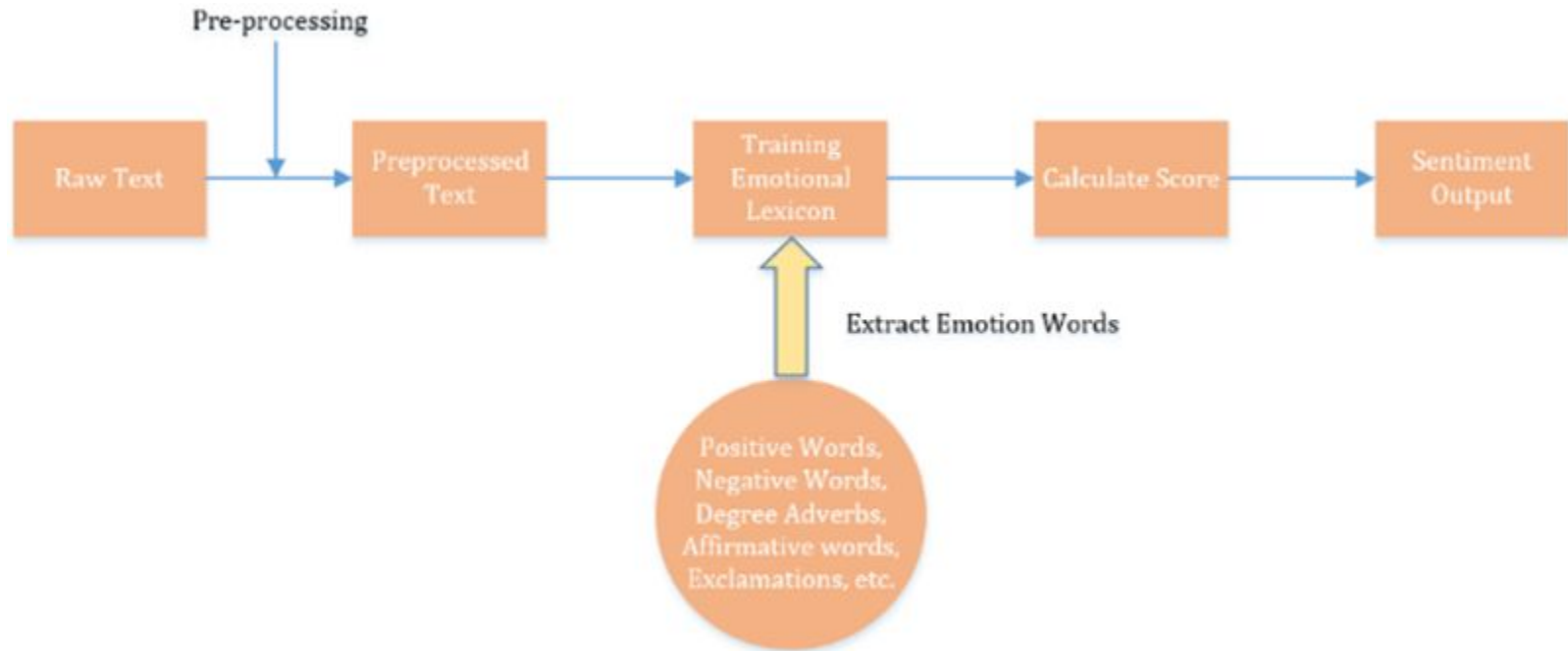
Preprocessing Mechanism

Sentiment Analysis



VADER'S Approach

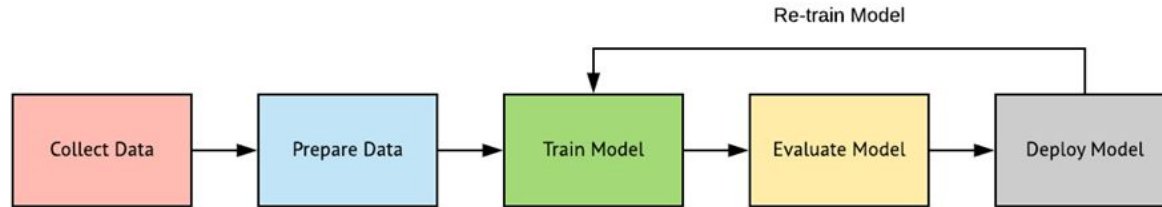
- The Valence Aware Dictionary and sEntiment Reasoner (VADER) is a rule-based sentiment analysis tool designed for efficient handling of social media text and short reviews. Therefore this tool is particularly effective for analyzing the user sentiments expressed in informal contexts, such as online product reviews.
- The sentiment scores are classified into four categories:
 - Positive
 - Negative
 - Neutral
 - Compound - a sum of the valence scores of each word ranging from -1 (most extreme negative) to +1 (most extreme positive)
- This scoring system provides accurate sentiment analysis, making it effective for texts written in informal language, for text with slang, abbreviations, emoticons, punctuation, and special characters.



The VADER method of referencing pre-defined lexicons

DistilBERT Approach

- Pre-trained DistilBERT multilingual base case model from HuggingFace
- Model was previously trained on 3000 samples
- The model was implemented as a pipeline
- Sentiment analysed as:
 - **Positive Sentiment:** Score ≥ 0.5
 - **Negative Sentiment:** Score < 0.5



Advantages of this approach:

- **Contextual Understanding:** DistilBERT's ability to understand context and relationships between words leads to highly accurate sentiment predictions, even in complex sentences.
- **Multilingual Capability:** The model's ability to handle multiple languages ensures that reviews in different languages are accurately analyzed.
- **Efficiency:** DistilBERT's optimized architecture allows for faster processing without significant loss of accuracy, making it suitable for large-scale sentiment analysis tasks.

The distilBERT model was trained by splitting the dataset into training and testing sets to evaluate the model's performance effectively.

DistilBERT Model Evaluation



The model was trained used Logistic Regression and the performance was plotted.

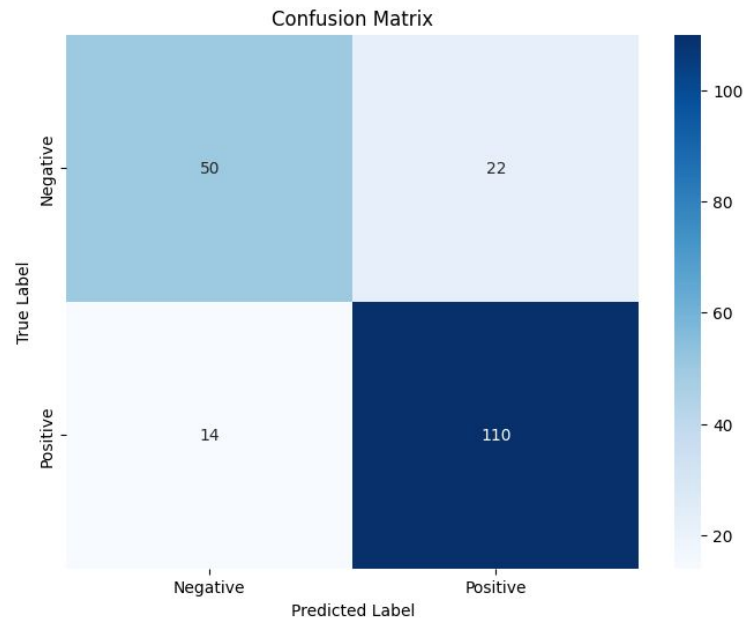
The average metrics were:

- Accuracy: 0.7785
- Precision: 0.8037
- Recall: 0.8375
- F1 score: 0.8194

DistilBERT Model Evaluation

A confusion matrix was plotted to find the number of False positives and False negatives

With further training, the number saw a gradual decrease from around 30-40 false predictions to 20s range with increase in model accuracy from 0.63 to 0.77.

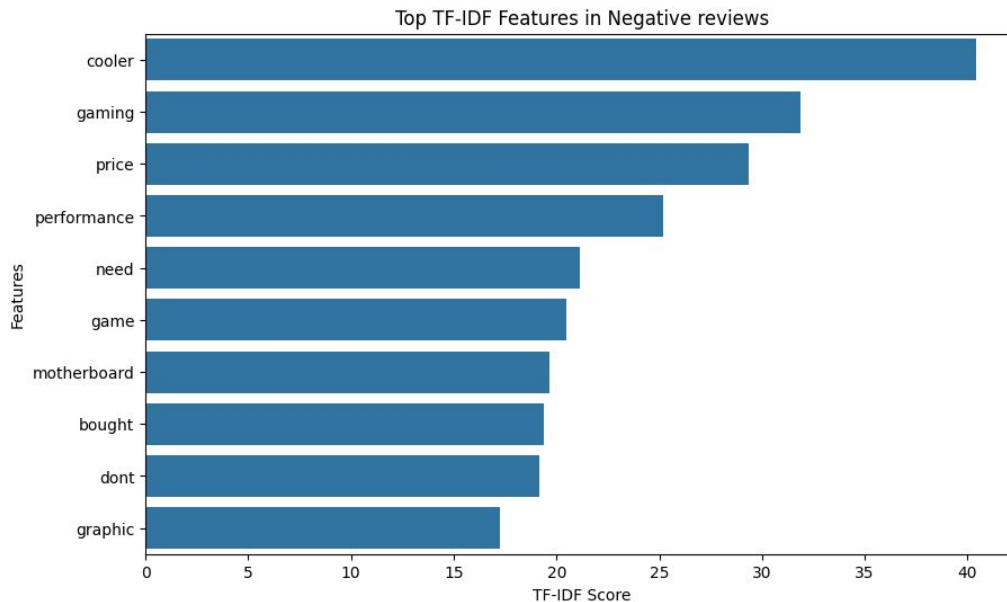


Feature Extraction

The Term Frequency-Inverse Document Frequency (TF-IDF) method was used.

This was done to identify the most common and significant words in both positive and negative reviews.

The top TF-IDF features were visualized using bar plots and word clouds.



Results and Discussion

13th generation processor comparison

Word Cloud of Positive reviews



VADER

Word Cloud of Positive reviews

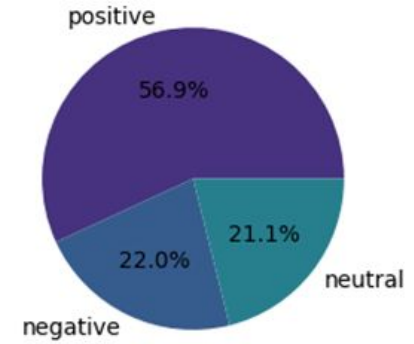
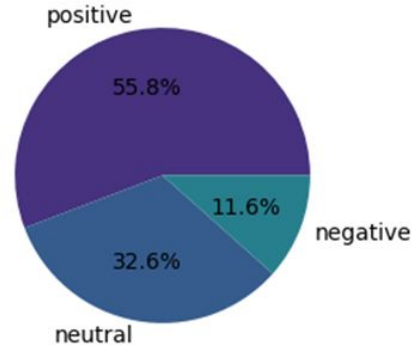
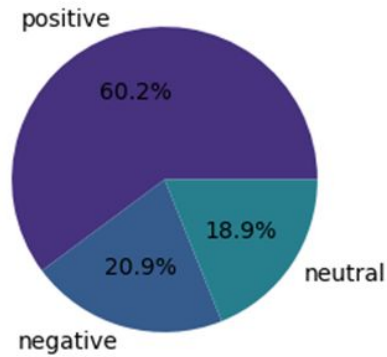


ML Pipeline

By using two analysis techniques, we were to arrive at a common result and get a more comprehensive report.

Generation	Positive Features	Negative Features
12th Gen	Price, Performance, Gaming, Good, Fast, Excellent	Cooler, Working, Broken, Motherboard
13th Gen	Price, Excellent, Performance, Recommend, Fast	Problem, Processor, Cooler, Damaged, Thermal, Temp
14th Gen	Power, Performance, BIOS Upgrade, RAM, AIO, Motherboard	Game Run, Work, Unstable, BIOS, Cooler Not

The analysis revealed the positive and negative sentiments across the different generations of Intel processor



These are distribution of user sentiments of different generations (12th, 13th and 14th respectively) showcasing the results of the sentiment analysis

Conclusion

- It is apparent that the 12th generation processors shows a relatively higher positive sentiment, with keywords such as "performance," "gaming," "good," "fast," and "excellent" being frequently mentioned. These terms show a strong user satisfaction related to performance and gaming capabilities.
- In contrast, the 13th generation reviews also highlight positive aspects like "price," "performance," and "fast," with a notable inclusion of "recommend," indicating strong consumer endorsement. However, this generation is marked by an increase in negative comments focusing on issues like "thermal temp" and "damaged," which reflect some concerns about reliability.
- The 14th generation introduces a broader range of positive keywords, including "power," "bios," "upgrade," and "ram," suggesting advancements and additional features. Despite these improvements, negative sentiment with mentions of "unstable," "bios," and "cooler," indicating ongoing issues with stability and hardware.

- The 14th generation shows significant improvements with the introduction of new positive keywords such as "power," "bios upgrade," and "ram," which were less emphasized in earlier generations. This suggests enhancements in hardware capabilities and feature sets. Despite the advancements, it continues to face issues with stability and reliability, as indicated by the increased negative comments.
- This study provided a comprehensive picture of the changing patterns in consumer satisfaction with Intel processors, highlighting the need of ongoing innovation and adaptability to market demands.
- This information is valuable for Intel to continue improving their products and addressing any recurring issues.

THANK YOU

Github: <https://github.com/Raaghashree/Intel-Products-Sentiment-Analysis>
Model: <https://huggingface.co/Raagha/finetuning-sentiment-model-3000-samples>