Problem Statement - 11
Intel Products Sentiment Analysis from Online reviews

Alex Stanley Alenchery - 2348307
Angeline A - 2348409
Aqshra Saji - 2348411
Gauri V Nair - 2348321
Vishnu E J - 2348466

Abstract

This independent study performs in-depth sentiment analysis of customer reviews across Intel's 11th, 12th, 13th, and 14th-generation processors using Amazon reviews dataset. Each processor variant was analyzed against a rich dataset of Amazon reviews to pinpoint the prevailing sentiments and critical insights. The results for the 11th and 12th generation processors were overwhelmingly positive, reflecting sentiments that highlighted performance, value for money, and gaming performance. However, recurring complaints cropped up regarding product delivery and the condition of the motherboards, thus pointing at possible improvements in user experience.

Moving into the 13th generation, comparative testing using models like Naive Bayes, SVM, Random Forest, and BERT proved the better performance of BERT with brilliant accuracy of 89.28%. Sentiment distribution shows a majority 65% positive reviews focusing on improved performance and energy efficiency. However, incompatibility issues and early technological glitches caused some negative sentiment. For the 14th generation, the multi-method approach using VADER—lexicon-based analysis, TextBlob—machine learning techniques, and LDA—advanced NLP model portrayed an overall positive reception. More specifically, improved performance, value for money offered, and thermal management were greatly praised, but high power consumption and poor compatibility showed lingering concerns. This piece of research contributes in part to the overall assessment of how effective state-of-the-art NLP methodologies are at extracting actionable insights from consumer feedback and guiding further product development or strategies toward customer satisfaction.

Introduction

Intel Corporation has innovated and delivered high-performance processors that power a vast range of computing devices. With each new generation of processors, Intel has pushed the boundaries of performance, efficiency, and integrated features, catering to the diverse needs of consumers, businesses, and tech enthusiasts. Analyzing how these products are perceived by users are crucial to maintain competitive edge and identify potential areas of improvement.

This project, conducted as part of the Intel Unnati Industrial Training Program 2024, performed a comprehensive sentiment analysis of i3, i5, i7, and i9 products of 11th Gen to 14th Gen processor reviews. By analyzing user reviews from e-commerce websites, we attempted to uncover trends and patterns in consumer sentiment, comparing the reception of different processor generations, and provide actionable insights that could help Intel refine its product offerings. We conducted analysis by grouping products of each generation to uncover the efficiency of processors with each successive generation. This analysis not only sheds light on consumer satisfaction but also helps in understanding the broader market trends and preferences. Through this sentiment analysis, we aim to contribute to Intel's mission of delivering exceptional computing experiences to its customers.

Literature Review

11th Generation Processors:

[1] SUMMARY: The 11th generation Intel processors codenamed Tiger Lake had notable progress in CPU architecture as well as integrated graphics. These CPUs were characterized by Willow Cove microarchitecture and Intel's 10nm SuperFin technology which greatly improved both performance as well as power efficiency.In simpler terms, some of the new upgrades in Tiger Lake included the following: Intel introduced Deep Learning Boost, and improved AI capabilities. This has really helped improve performance in AI tasks and workloads. Key features included Intel's new Xe graphics architecture that significantly boosted integrated graphics performance as well as Thunderbolt 4 and PCIe 4.0 support.

GAPS: Although significant improvement was realized in graphics and AI performance by Tiger Lake, significant enhancements might as well be further explored by undertaking more research to determine application areas in real life such as Thermal management and efficiency during an even heavier working content.

[2] SUMMARY: Intel's newest processors to be released in 2021 exhibit significant architectural advancements and highlight improvements in artificial intelligence. In order to support efficiency, more focus has been placed on boosting throughput for each unit of power (watt) consumed among other features like optimizations towards achieving high efficiency ratios through re-engineering various systems so that they can better handle multiple tasks concurrently. The paper explains how these developments help to enhance real-world efficiency and performance in domains like industrial automation systems and healthcare technologies.

[3] SUMMARY: The current research project examines the enhancements in heat dissipation and power handling in the 11th generation processors made by Intel. It also identifies and proposes methods that can be employed to reduce power consumption and improve heat transfer in these devices so as to evade difficulties linked with increased effectiveness and reduced dimensions. A detailed look at the thermal management schemes subsequently reveals their overall effect on processor life span as well as its dependability that is given by the authors.

12th Generation Processors:

[4] SUMMARY: Intel Alder Lake, the 12th generation of processors, brought up an architecture where it's called hybrid (mixing high-performance (P-cores) with high-efficiency cores (E-cores)). In general terms, this was done in order to get better performance from multi-threaded tasks and also save power. The innovations were Intel Thread Director allowing for a more even distribution of workloads across the cores, an enhanced Intel Iris Xe GPU that has stronger graphic capabilities than any other integrated one so far, as well as support of new standards such as PCIe 5.0 and DDR5 memory.The processors are suitable for applications ranging from high-performance computing (HPC) to embedded systems. The hybrid architecture enhances AI performance while also increasing energy efficiency, which can be immensely beneficial if you are planning on using it even during power outages.

GAPS: Although there were hopeful enhancements seen with the integration model; realistic tests are still necessary for the size of these systems as they are in use today and the advancement of more effective artificial intelligence and stronger security features in the future.

[5] SUMMARY: At the moment, Intel's Raptor lake processors have a dual upper hand over AMD's Ryzen 7000 series in that they are faster and cheaper relative to performance. This therefore gives Intel a good position especially in segments such as core i3 and i5 ranges in quotation. Nonetheless, this could change considering that Intel may lower its prices to shift the competition dynamics in cheaper categories favoring Amd.

GAPS: The talk touches on the "in-the-now" advantage of Intel's Raptor Lake when it comes to performance and price, though we don't know for sure if it is going to be there next year or the year after. More qualitative studies should try to investigate how different pricing strategies have influenced companies sustainability' in terms of competitiveness over time. For example, one still does not understand how changing prices can influence expectations of consumers and at the same time affect market shares dynamics.

13th Generation Processors:

[6] SUMMARY: Intel's new Raptor Lake-S processors assure market end-users with a choice of many different SKUs. The family includes the performance-driven Core i9 and Core i7 models, each with a 65 W TDP, which can be the reasons for having impressive multi-threaded performance and notable clock speeds. The Core i5 line is split into multiple models, the introduction of which is the ability to configure the performance and power efficiency. On the other hand, the Core i3 series contains relatively cheap quad-core processors, but they are without efficient cores. These processors work with DDR5-5600 and DDR4-3200 memory which corresponds to various workloads such as gaming and everyday computing.

GAPS: Intel's Raptor Lake-S processors fall across a full range of power and functionality, while there is a need to fill up some knowledge gaps which are concerned with extensive experimentation into the scalability of the hybrid architecture, particularly in real-world performance situations. Furthermore, a well-researched idea of thermal management and effectiveness per various workloads increases the potential for the enhancement of these processors' applications. The exploration of the influence of memory support, in particular, DDR5-5600, on system-wide performance and user experience is the crucial point of the issue.

14th Generation Processors:

[7] SUMMARY: A large lineup of SKUs is now available from Intel with the introduction of their 14th Gen Core series of processors, addressing a variety of user needs. Some of their highlighting products are like the K-series flagship chips such as Core i9-14900 that demand for the best performance via a 125 TDP (and maybe even a 219W TDP going forward), as well as the non-K options with lower 65W and 35W TDPs for use in size-constrained or efficiency-focused designs. Intel has expanded its range with recent additions; things like the Core i3 series when it lacks efficiency cores, and some Intel Processor 300 models for the lower-end. With these chips supporting both DDR5 and DDR4 memory, users have more choice in the platform they pursue. IPC and power efficiency are improved across the lineup according to Intel, putting these CPUs on an even playing field with AMD— however, testing is required to prove the performance of these chips.

GAPS: Research indicates a dearth in understanding when it comes to performance audits that have not been conducted in an industry-controlled environment, as well as few tests of the linearity of AMD system frequencies when workloads change. The new infrastructure could be another path to test power efficiency for better thermal management once we have a full view of it. Likewise, it would be wise to test how support for DDR5 and DDR4 memory affects overall system performance before anyone else has that answer; such analysis would be greatly useful for those who need to make an informed decision about the two's relative cost and compatibility.

Data Collection

Data sources: Amazon user reviews were used as a source, with an emphasis on products associated with different Intel processor generations and models. Because of its large user review collection, which offers a rich dataset for sentiment analysis, Amazon was selected.

Data acquisitions: The information was gathered by use of web scraping methods. In particular, Google Chrome extensions made for data extraction from websites were used to scrape reviews. The gathering of review texts, ratings, dates, and other pertinent data was made easier by these extensions. After that, the data was scraped and exported in CSV format for additional examination.

Data descriptions: The sentiment analysis dataset consisted of 3,234 comprehensive reviews of different generations and models of Intel processors. These evaluations, which were taken from Amazon, represented a wide variety of consumer opinions and experiences. Every review featured extensive user discussion written in text, along with ratings and, occasionally, visual material such as product photos. This large dataset allowed for a detailed investigation of consumer attitudes and views regarding Intel processors, providing information on user preferences, performance assessments, and product satisfaction. The evaluations, which covered a wide range of time periods, offered a strong basis for in-depth sentiment analysis, exposing patterns and trends in customers' perceptions over time.

Data Preprocessing

- *11th Gen*
    - To guarantee that the reviews were appropriately organized and processed, the sentiment analysis procedure for the 11th generation Intel processors required a number of vital data pretreatment steps:
        1. **Data Loading:** Imported the customer review data from CSV files into a pandas DataFrame. Ensured that the review text column was correctly formatted as strings to facilitate further text processing.
        2. **Text Tokenization:** Tokenized the review texts using the RobertaTokenizer from the Hugging Face Transformers library. In this phase, the text was formatted to fit the RoBERTa model, which included padding and truncating it to a consistent length.
        3. **Model Initialisation:** loaded the pre-trained RoBERTa model for sequence classification (cardiffnlp/twitter-roberta-base-sentiment). The efficacy of this model in sentiment analysis tasks led to its selection.

4. **Tokenization Function:** Using the tokenizer, a function was built to preprocess every review text, guaranteeing uniformity and effectiveness throughout the dataset's processing.
5. **Sentiment Prediction and Mapping:** developed a function that uses the RoBERTa model to predict each tokenized review's sentiment. The model produces logits, which torch.argmax was used to transform to sentiment classes (negative, neutral, and positive). For easier interpretation, the numerical sentiment predictions were mapped to the descriptive labels (positive, neutral, and negative).
6. **Data Exporting:** The processed DataFrame was saved to a new CSV file for additional reporting and analysis. It contained the original text, numerical sentiment ratings, and descriptive sentiment labels.

- *12th Gen*
  - **Data Loading**: Imported the customer review data from a CSV file to a pandas DataFrame. Ensured that the review text column was correctly formatted as strings to facilitate further text processing.
  - **Text Tokenization**: Utilized the RobertaTokenizer from the Hugging Face Transformers library to tokenize the review texts. This step involved converting the text into a format suitable for the RoBERTa model, including padding and truncating the text to a consistent length.
  - **Model Initialization**: Loaded the pre-trained RoBERTa model (cardiffnlp/twitter-roberta-base) for sequence classification. This model was chosen for its effectiveness in sentiment analysis tasks.
  - **Tokenization Function**: Defined a function to preprocess each review text using the tokenizer, ensuring consistency and efficiency in processing the entire dataset.
  - **Sentiment Prediction**: Created a function to predict the sentiment of each tokenized review by passing it through the RoBERTa model. The model outputs logits, which are then converted to sentiment classes (negative, neutral, positive) using torch.argmax.
  - **Mapping Sentiment Labels**: Mapped the numeric sentiment predictions to descriptive labels (negative, neutral, positive) for better interpretability.
  - **Adding Sentiment Columns**: Added two new columns to the DataFrame: one for the numeric sentiment score and another for the descriptive sentiment label, making the results easily accessible and understandable.
  - **Data Export**: Saved the processed DataFrame, which included the original text, numeric sentiment scores, and descriptive sentiment labels, to a new CSV file for further analysis and reporting.

- *13th Gen*
  - **Date Extraction and Conversion**: We extracted and converted all date information into a standard format, hence helping in chronological analysis and time-based visualizations. Sometimes, even this could be quite important for consistency, hence making it easier to manipulate the data regarding dates by converting the date entries into one standard datetime format.
  - **Language Detection and Translation**: In order to standardize the language for consistent sentiment analysis, language detection and translation are done. The langdetect library detects the language of each entry in text, while the googletrans library does the translation from non-English text to English. All the text data will then be in a standard language, thus enabling sentiment analysis to be homogeneous and reliable with TextBlob providing the sentiment polarity scores for the translated text.
  - **Text Cleaning**: Text cleaning removes noise and standardizes text data to make it ready for any analysis. This is constituted by the removal of special characters and punctuation using regular expressions, case conversion of text to lower case in order to avoid case sensitivity, tokenization of the text into words, and the removal of common stop words based on a predetermined list from NLTK. This will make the text very clear, consistent, and focused on meaningful content to help improve the accuracy of sentiment analysis.
- *14th Gen*
  - **Date Handling**: Extracted dates from review text strings and converted them into a standardized datetime format, facilitating temporal analysis and trend identification.
  - **Language Detection and Translation**: Applied the `langdetect` library to determine the language of each review. Reviews in languages other than English were translated to English using the Google Translate API, ensuring consistency in sentiment analysis.
  - **Text Cleaning**: Performed several text cleaning steps to prepare the data for analysis:
    1. *Special Characters Removal*: Stripped out special characters and punctuation to maintain focus on meaningful text content.
    2. *Lowercasing*: Converted all text to lowercase to ensure uniformity.
    3. Stop Words Removal: Utilized NLTK to remove common stop words that do not contribute to sentiment.
    4. *Tokenization*: Split the text into individual tokens for further processing.
    5. *English Word Filtering*: Checked the proportion of English words to ensure that non-English texts were accurately identified and translated if necessary.

- **Text Normalization**: Applied tokenization and lemmatization to reduce words to their base forms, aiding in more accurate sentiment analysis.
- **Sentiment Analysis Preparation**: Created a clean and structured dataset that included original text, cleaned text, detected language, and translated text, setting the stage for detailed sentiment scoring and analysis.

Sentiment Analysis Methodology

We employed different methodologies for sentiment analysis across each generation which are detailed in this section. Each generation's sentiment analysis utilized different approaches, models, and feature extraction to compare and understand the working and efficiency in understanding consumer sentiments.

- *11th Gen*
  - **Approach:** The workflow involves using the DistilBERT model, a pre-trained transformer model from Hugging Face, specifically fine-tuned for sentiment classification tasks. First, we import the required libraries. Then, the procedure begins by identifying and loading all CSV files from the directory into a single dataframe. The text data is prepared by tokenization using the DistilBERT tokenizer to transform it into a format suitable for model input, and thereby loads the DistilBERT model, which has been fine-tuned for sentiment analysis.The sentiment of each text entry is predicted using the loaded model. Finally, the DataFrame, now including sentiment predictions, is saved to a new CSV file.
  - **Model Selection:** Models for Intel 11th Gen CPU sentiment analysis were chosen based on their accuracy and capacity to manage the intricacies of textual input. DistilBERT, Naive Bayes, SVM, and Random Forest are among the models selected; each has particular advantages in processing and interpreting sentiment. DistilBERT was selected due to its exceptional deep learning performance, attaining the most precision in sentiment classification. It was the perfect option for encapsulating the opinions voiced regarding Intel 11th Gen processors because of its capacity to comprehend contextual subtleties in language. Because of its dependability and simplicity, Naive Bayes was chosen. Even though it was a more basic model, it performed well and had an accuracy that was on par with DistilBERT, which made it a useful tool for preliminary sentiment extraction and analysis. **Naive Bayes** was selected for its simplicity and reliability, despite being a simpler model. Support Vector Machine, SVM, has an accuracy of 0.73 and a robust performance in handling high-dimensional data. Random forest and RoBERTa are an added contribution with a better stability and providing a balanced view towards sentiment trends, especially in detailed reviews and discussions about products. By combining these models, a comprehensive

approach to sentiment analysis was achieved, collecting different facets of textual nuances and offering trustworthy insights into the attitudes surrounding Intel 11th Gen processors.

- **Feature Extraction:** To guarantee reliable sentiment analysis of Intel 11th Gen processors, the models were assessed using accuracy, precision, recall, and F1 score. The percentage of accurate classifications is represented by accuracy, which provides a general indicator of the model's effectiveness. By calculating the percentage of real positive predictions among all positive predictions, precision evaluates the accuracy of the discovered positive feelings. Recall gauges how well the model captures real positive examples, demonstrating how thorough sentiment recognition is. In order to ensure accurate and dependable sentiment analysis, the F1 score integrates accuracy and recall into a single statistic that balances the model's capacity to identify true positives and prevent false negatives.

- *12th Gen*
  - **Approach:** For this sentiment analysis, a deep learning approach was used for its enhanced ability to capture complex patterns and nuances in data.
  - **Model Selection:** We selected the RoBERTa model to do the sequence classification task, an advanced transformer-based model that has been pre-trained on a large variety of text, making it highly effective for text classification. The specific model applied here was 'cardiffnlp/twitter-roberta-base-sentiment', which is fine-tuned for sentiment analysis tasks.
  - **Feature Extraction:** For feature extraction, the Roberta Tokenizer was used. This preprocesses the text data by adding padding and truncating it to ensure uniform input size. It then converts the text data into a tensor format that is appropriate for the model.

- *13th Gen*
  - **Approach:** Initially a rule-based system with predefined lexicons and heuristic rules performed basic text preprocessing. Sentiment extraction was also executed. This foundational layer enabled quick identification and handling of common textual patterns. To enhance analysis we incorporated machine learning algorithms such as Naive Bayes. Support Vector Machine (SVM) and Random Forest were also used. Naive Bayes, known for its simplicity and efficiency with text data was complemented by SVM's capability. It handles high-dimensional spaces. Random Forest's ensemble learning approach was beneficial. For capturing complex text patterns, deep learning models like BERT and RoBERTa were utilized. BERT is renowned for its contextual understanding. It provided advanced sentiment classification capabilities.

- **Model Selection:** Model selection was based on accuracy and ability to handle textual intricacies. BERT and Naive Bayes achieved the highest accuracy of 0.80. BERT was chosen for its deep learning prowess. Naive Bayes was selected for its simplicity. And reliability SVM with an accuracy of 0.73 was selected for its robust performance in high-dimensional text data. Random Forest achieving 0.66 accuracy was included for its ensemble approach. This added diversity.
- **Feature Selection:** The model has been assessed in terms of accuracy, precision, recall, and F1 score. The definition of accuracy lies in percentage of accurate classifications. The measure of how many true positive predictions were there among all positive forecasts is called precision. Recall shows what part of the correct positive instances was taken into account. The F1 score combines precision and recall in terms of false positives and false negatives under one formula – thus detecting these mistakes equally well.

- 14th Gen
  - **Approach**: The sentiment analysis for the 14th generation Intel processors employed a comprehensive approach combining multiple techniques. The process began with data collection from various customer reviews. The analysis utilized natural language processing (NLP) techniques, including text preprocessing, tokenization, and the removal of stop words. Both lexicon-based methods and machine learning algorithms were applied to capture the nuances of customer sentiment.
  - **Model Selection**: Several models were used in the analysis to ensure a robust understanding of sentiment:
    - VADER (Valence Aware Dictionary and sEntiment Reasoner): This rule-based sentiment analysis tool was used for its ability to handle sentiments expressed in social media contexts.
    - TextBlob: Employed for its simplicity and effectiveness in providing polarity and subjectivity scores.
    - Latent Dirichlet Allocation (LDA): Used for topic modeling to identify key themes in the reviews.
    - TF-IDF (Term Frequency-Inverse Document Frequency): Applied to understand the importance of words in the document corpus.
  - **Feature Extraction**: The analysis extracted various features to capture sentiment:
    - Polarity scores: Measuring the positive, negative, and neutral sentiment of each review.
    - Compound scores: Providing a single sentiment measure.
    - Aspect-based sentiment: Analyzing sentiment towards specific aspects like performance, price, power consumption, and temperature.

- ◇ Bigrams: Identifying frequently co-occurring word pairs in positive and negative reviews.
- ◇ Language detection: To understand sentiment across different regions and languages.

Implementation

The implementation of the sentiment analysis for the Intel processors involved several key steps to ensure accurate and insightful results. This process included selecting appropriate tools and libraries, preparing the data, and training the sentiment analysis model. The implementation for each generation is given below.

- *11th Gen*
    - **Tools and Libraries:** The following libraries and tools were utilized to apply sentiment analysis for the 11th generation product reviews:
        1. Python: The primary programming language used for implementing the sentiment analysis pipeline.
        2. Transformers: A library by Hugging Face that provides pre-trained transformer models and tokenizers. Specifically, transformers for accessing RoBERTa models and tokenizers.
        3. PyTorch: An open-source machine learning library used for model training and evaluation. PyTorch provides the underlying framework for the RoBERTa model's operations and tensor computations.
        4. Pandas: A library used for data manipulation and analysis. It was employed for handling and processing the dataset of customer reviews.
        5. NLTK (Natural Language Toolkit): For NLP tasks such as tokenization and stop word removal.
    - **Model Training:** The following steps were involved in using and training the RoBERTa model:
        1. Loading Data: The cleaned dataset containing user reviews was loaded using pandas.
        2. Initializing the Tokenizer: The RobertaTokenizer was initialized to preprocess the text data.
        3. Preprocessing and Tokenizing Text: A function was defined to preprocess and tokenize the text data, ensuring that the text was appropriately padded and truncated to fit the model’s input requirements.

4. Loading the Pre-trained Model: Utilized the pre-trained RoBERTa model fine-tuned for sentiment analysis
5. Predicting Sentiment and Mapping to Labels: A function was created to perform sentiment prediction, and the predicted classes were mapped to human-readable sentiment labels (Negative, Neutral, Positive) using a predefined dictionary.

- *12th Gen*
  - **Tools and Libraries**: To implement the sentiment analysis for 12th generation product reviews, the following tools and libraries were used:
    1. Python: The primary programming language used for implementing the sentiment analysis pipeline.
    2. Transformers: A library by Hugging Face that provides pre-trained transformer models and tokenizers. Specifically, transformers for accessing RoBERTa models and tokenizers.
    3. PyTorch: An open-source machine learning library used for model training and evaluation. PyTorch provides the underlying framework for the RoBERTa model's operations and tensor computations.
    4. Pandas: A library used for data manipulation and analysis. It was employed for handling and processing the dataset of customer reviews.
    5. Torch: PyTorch's core library, utilized for tensor operations and model inference.
  - **Model Training:** The process of training and utilizing the RoBERTa model included the following:
    1. Loading Data: Cleaned dataset containing user reviews were loaded using pandas.
    2. Initializing the Tokenizer: The RobertaTokenizer was initialized to preprocess the text data.
    3. Preprocessing and Tokenizing text: We defined a function to preprocess and tokenize the text data.This function ensures that the text is appropriately padded and truncated to fit the model’s input requirements.
    4. Loading the pre-trained model: Utilized the pre-trained RoBERTa model fine-tuned for sentiment analysis. The specific model used was 'cardiffnlp/twitter-roberta-base-sentiment', which was loaded using the Transformers library.
    5. Predicting Sentiment and Mapping to labels: A function was created to perform sentiment prediction, and the predicted classes

were mapped to human-readable sentiment labels (Negative, Neutral, Positive) using a predefined dictionary.

- *13th Gen:*
    - **Tools and Libraries:** We used a range of software tools and libraries to ensure efficient data handling, model training, and evaluation. The primary tools and libraries used include:
        1. Python: The main programming language for scripting and implementation.
        2. Pandas: For data manipulation and analysis.
        3. NumPy: For numerical computations and array operations.
        4. NLTK: For NLP tasks such as tokenization and stop word removal.
        5. scikit-learn: For machine learning algorithms, model training, and evaluation.
        6. TensorFlow and Keras: For building and training deep learning models like BERT.
        7. Transformers: For leveraging pre-trained models like BERT and RoBERTa.
        8. Matplotlib and Plotly: For data visualization and plotting.
        9. Jupyter Notebook: For interactive coding and documentation.

    - **Model Training:** The data set was divided into training (70%), validation (15%) and test (15%) respectively. Tuning of hyperparameters was done for every model: Naive Bayes, SVM, Random Forest and BERT. Training took different durations where Naive Bayes was completed in minutes and while more complex models like BERT took several hours.
    - **Evaluation Metrics:** We evaluated the performance of the model using different metrics such as accuracy, precision, recall, and F1 score. Overall correctness was measured by accuracy, positive predictions accuracy by precision, while recall measured the ability to identify all positive instances, and F1 score provided a balanced measure of precision and recall. The choice of the most useful model sentiment analysis depended on these metrics.

- *14th Gen:*

    - **Tools and Libraries:** For the sentiment analysis of the 14th generation Intel processors, we used a variety of tools and libraries to ensure robust data handling, analysis, and visualization. The primary tools and libraries used include:

1. Python: The main programming language for scripting and implementation.
2. Pandas: For data manipulation and analysis.
3. NumPy: For numerical computations and array operations.
4. NLTK (Natural Language Toolkit): For NLP tasks such as tokenization and stop word removal.
5. TextBlob: For basic sentiment analysis.
6. Langdetect: For language detection.
7. Googletrans: For translating non-English reviews to English.
8. Scikit-learn: For feature extraction and machine learning tasks.
9. VADER (Valence Aware Dictionary and sEntiment Reasoner): A pre-trained model used for determining the sentiment of the text.
10. Matplotlib and Seaborn: For data visualization and plotting.
11. WordCloud: For visualizing common words in reviews.
12. Statsmodels: For time series analysis.

- **Pre-trained Model: VADER Sentiment Analysis:**
    1. The VADER sentiment analysis model was employed to determine the sentiments expressed in the reviews. VADER is particularly suited for social media texts and can handle slang and emoticons, making it a good choice for analyzing customer reviews.
- **Analysis Methods:**
    1. Temporal Analysis: Seasonal decomposition was performed using the seasonal_decompose function from Statsmodels to identify trends and patterns in the sentiment over time.
    2. Language-specific Sentiment Trends: Sentiment scores were analyzed across different languages to understand regional differences in customer perceptions.
    3. Common Themes Identification: Using CountVectorizer and Latent Dirichlet Allocation (LDA), Common themes and topics were identified in both positive and negative reviews. Word clouds were also generated to visualize the most frequent terms in the reviews.

Results and Discussion

- *11th Gen:*
  - **Model Performance:** A range of machine learning strategies and natural language processing models were used in the sentiment analysis for Intel processors, namely the i5-11600K, i5-11400F, i9-11900K, i7-11700, i5-11400, and i7-11700K models. Because advanced models like RoBERTaForSequenceClassification and BERT have a high degree of accuracy in interpreting and categorizing textual sentiment, they were employed. These models, refined on a variety of datasets, are excellent at interpreting complex emotions found in customer evaluations.

    A high degree of consumer satisfaction with these processors is indicated by the sentiment distribution across the 703 customer reviews. Although there are significantly fewer good reviews than negative ones, the majority of them are positive.
  - **Sentiment Distribution:** The distribution of sentiments across the 703 customer reviews is as follows:

    Ratings:
    - 5-star ratings: 523
    - 4-star ratings: 94
    - 3-star ratings: 41
    - 2-star ratings: 10
    - 1-star ratings: 35

    Sentiments:
    - Positive: 350
    - Negative: 353

    The majority of the evaluations are favorable, suggesting that these processors are well-liked by customers. Less often occurring neutral and negative attitudes imply that significant unhappiness is not pervasive.
  - **Insights:**
    - Positive Reviews:
      1. Improved Performance: Numerous people praised the improved performance of the CPUs, pointing out faster speeds and improved multitasking skills. Positive remarks emphasized how well the processors performed in resource-intensive tasks such as video editing, gaming, and 3D rendering.
      2. Energy Efficiency: Numerous evaluations highlighted the processors' energy efficiency, pointing out that they consumed less power than models from earlier generations.

3. Value for Money: Consumers believed that the CPUs provided outstanding performance at a reasonable cost.
4. Advancements: Positive sentiments were associated with the processors' cutting-edge features, such as AI capabilities, integrated graphics, and support for the latest connectivity standards like PCIe 5.0 and DDR5 memory.

◇ Negative Reviews:

1. Compatibility Issues: Certain customers experienced issues with motherboard compatibility, necessitating BIOS updates and leading to extra expenses and aggravation.
2. Price Concerns: A few reviews brought up how much these CPUs cost more than comparable goods from rival manufacturers, especially AMD.
3. Performance Inconsistencies: A few users reported experiencing extreme heat and slowness when working with heavy workloads.
4. Initial Bugs and Glitches: Early adopters complained of instability, system crashes, and firmware faults; however, these problems were usually fixed with upgrades.

- **Discussion:** Customer evaluations' sentiment analysis indicates that Intel processors are typically well-liked. Customers praise the items for their more affordable prices, better performance, and increased energy efficiency. Compatibility problems and cost concerns, however, are major points of contention. Resolving these problems could improve customer satisfaction even more.

  Intel gathers insights into consumer impressions with sophisticated natural language processing (NLP) techniques including sentiment scoring and topic modeling. These findings inform product development and customer support initiatives.

- *12th Gen:*
  - **Model Performance:** The selection of the RobertaForSequenceClassification model, specifically the 'cardiffnlp/twitter-roberta-base-sentiment', for sentiment classification was driven by its advanced capabilities in understanding and classifying textual sentiment. RoBERTa, a robustly optimized variant of BERT, has demonstrated exceptional performance in various natural language processing tasks, including sentiment analysis. The 'cardiffnlp/twitter-roberta-base-sentiment' model is fine-tuned on a diverse dataset of tweets, making it particularly adept at handling nuanced sentiment expressions in text. This pre-trained model's ability to capture subtle emotional tones and its high accuracy in sentiment classification

make it an ideal choice for analyzing the sentiment of customer reviews, ensuring reliable and insightful results.

- **Sentiment Distribution:** There are 763 5-star ratings, 144 4-star ratings, 44 3-star ratings, 22 2-star ratings, and 77 1-star ratings. Most of the sentiments are positive, while neutral sentiments are far more than the negative. Extreme negative sentiments are notably rare, suggesting that while dissatisfaction exists, it is not overwhelmingly prevalent. The sentiments were positive in the year 2021 and saw a decline in 2022-2023 and an increase again in the year 2024.
- **Insights:** The sentiment analysis and visualizations reveal several key insights into customer reviews for the 12th generation processors The violin plot revealed that positive sentiments are heavily skewed towards higher ratings, indicating that satisfied users give higher ratings. The word cloud highlights prominent themes such as "CPU," "Processor," "Price," and "Gaming," which dominate user reviews. These keywords suggest that discussions frequently center around product performance, Average sentiment score increased towards the year 2023-2024, this may indicate that recent improvements and new processor releases may be contributing to a more favorable reception in the latest years. In contrast, negative sentiment reviews often feature words like "cpu," "motherboard," "cooler," and "arrived." Notably, "arrived" and "motherboard" appear prominently, indicating dissatisfaction with product arrival and the condition of the motherboard. This suggests that unboxing experiences hold a significant value in customer sentiments. The bar chart analysis of negative sentiments identifies "faulty motherboard" and issues with "packaging" as prominent concerns. These issues reflect a broader dissatisfaction with product condition upon arrival, as highlighted by the frequent negative mentions related to packaging and the product's initial state.
- **Discussion:** The sentiment analysis of customer reviews for Intel processors reveals a complex landscape of user satisfaction and dissatisfaction. The distribution of sentiment polarity, ranging from -1 to 1, shows that neutral sentiments are the most common, with a notable prevalence of polarity values around 0. This indicates that while a significant portion of reviews does not exhibit strong emotions, there is a healthy distribution of opinions across the spectrum. Conversely, negative sentiment reviews highlight issues with terms such as "cpu," "motherboard," "cooler," and "arrived." The recurrence of "arrived" and "motherboard" in negative reviews points to significant concerns regarding product condition upon arrival and the quality of the motherboard. These issues suggest that customers are particularly dissatisfied with their unboxing experience and the initial state of the product, which is crucial for overall satisfaction.This indicates that improvements should be made not only on the specifications of the processor but also on the packaging. In summary, while overall sentiment towards

Intel processors tends to be positive, specific areas of dissatisfaction, particularly related to product condition and packaging, are evident. The analysis indicates that addressing these concerns could enhance user satisfaction significantly. Future research should delve deeper into individual product reviews and consider the impact of ongoing product improvements and packaging enhancements to validate and address these issues.

- *13th Gen:*
  - **Model Performance:** The project on sentiment analysis analyzed different models that can be used to tell how Intel's 13th generation processors are being viewed by looking at Amazon reviews. When it comes to performance, we used accuracy, precision and recall as well as F1 score across dissimilar models such as Naive Bayes, SVM, Random Forest and BERT amongst others, whose measures were calculated using varying parameters. Having 89.28% BERT had the highest percentage while SVM stands at 79%, Gradient Boosting - 76.5%, Random Forest and Naive Bayes had the same performance in the13th Gen dataset – 75.5% and LSTM had the very least accuracy for the 13th Gen intel processor reviews dataset. These trends were the same for the precision as well as recall scores. They are consistent among BERT algorithms that consistently outperform in identifying positive, negative or neutral sentiments in reviews.
  - **Sentiment Distribution:** Patterns that were enlightening within the dataset were brought to light as the analysis of sentiment distribution was carried out. 65% of these reviews were characterized by positive sentiments, whereas only 25% contained negative ones and the rest had neutral sentiments. This implies that Intel's 13th generation processors have generally been well received by the majority of reviewers although some have apprehensions or complaints.
  - **Insights:**
    - ◇ **POSITIVE REVIEWS**
      1. **Performance Improvements:** Many of the reviewers from amazon were happy about the enhancements in the 13th Gen Processors. The Users most often mentioned faster processing speeds in the reviews, also mentioned about multitasking capabilities of the 13th Gen processors, and it has better overall system responsiveness. The users appreciated the Gaming applications and resource intensive applications such as video editing, 3D rendering.
      2. **Energy Efficiency:** There were so many positive reviews about the energy efficiency of the 13th Gen processors. User mentioned lower power consumption noticed in 13th gen compared to previous Generations.

3. **Value for Money:** Many users who reviewed felt that the 13th generation processors offered excellent value for money. They concluded that the 13th gen processors had superior performance when compared to the previous Generation at the price point.
4. **Technological Advancements:** The 13th Gen processors were highly rated by the reviewers for their technological advancements like better AI capabilities, integrated graphics(IGP), improved support for the latest connectivity standards among others which include PCIe 5.0 and DDR5 memory. Hence, such aspects basically make these CPUs to be platforms that are primed at least in the near term future making them alluring when thinking in terms of using them for extended periods.

◇ **NEGATIVE REVIEWS**

1. **Compatibility Issues:** Many negative reviews revolve around compatibility issues. While trying to connect 13th Gen processors to the current hardware or software configurations, problems were met by users. Consumers lodged several complaints including bios updates, motherboard compatibility and driver support, which resulted in extra expenses as well as increased dissatisfaction among upgrading individuals.
2. **Price relative to Competitors:** While most of the reviewers thought the 13th gen processors from Intel gave a good bargain, there were some of the users who found the pricing still on the higher side especially when compared with other competing products in the market with the same capabilities. In particular, such clients thought AMD chipsets offered better bang for buck at specific price levels thus driving their negativity towards them.
3. **Performance Inconsistencies:** Certain workloads or applications caused some users to experience inconsistency in performance according to them. Under these performance reviews it was narrated how the continued use under heavy loads resulted in unsatisfactory slowness and high-temperature though the 13th Gen processors had a lot more good reviews
4. **Initial Bugs and Glitches:** Negative reviews indicated a system that consistently crashes, unstable system and bugs in firmware were cited as issues by the early users. Nevertheless, these challenges marked a bad beginning for early users of this processor type before being solved through updates.

- **Discussion:** Amazon reviews are benefiting from machine learning (ML) as well as deep learning (DL) Sentiment analysis by BERT received a high accuracy ranking at 89.28%,. This indicates that there were many more happy than sad reactions towards Intel's 13th Gen CPUs due to how they functioned and the cost at which they sold. What we can learn from sentiment data is useful information for Intel: it shows their weaknesses and their strengths. This means that there are ways that they can make better products or help people who feel bad about them fro reviewing poorly on them so as to attain more consumer loyalty, re-buying products again and again even after an initial purchase was a recommendation made by most clients during the focus group session convened by Intel last week and it seems indeed that there might be an opportunity here which is worth exploring whether we should give this strategy some serious thought while working on our strategy. Overall, the sentiment analysis methodology does more than enabling you grasp the attitude of customers in totality but also goes an extra mile in revealing the impact of advanced NLP approaches on provision of implementable opinions from broad text collections that help determine decisions made by companies that want to remain competitive by coming up with new products or strategies for convincing customers to continue buying them.

- *14th Gen:*
  - **Model Performance**: The sentiment analysis for Intel's 14th generation processors utilized various models and techniques to analyze customer reviews. While specific accuracy metrics for individual models weren't provided, the analysis employed a combination of lexicon-based methods (VADER), machine learning techniques (TextBlob), and advanced NLP models (LDA for topic modeling). This multi-faceted approach allowed for a comprehensive understanding of sentiment across different dimensions.
  - **Sentiment Distribution**: The analysis revealed a predominantly positive sentiment towards the 14th generation Intel processors. Out of 502 reviews analyzed, 350 (69.7%) were categorized as positive, 99 (19.7%) as negative, and 53 (10.6%) as neutral. This distribution indicates a generally favorable reception of the new processors among consumers.
  - **Insights**:
    - **Positive reviews-**
      1. Performance Improvements: Many reviewers praised the enhanced performance of the 14th Gen processors. Users frequently mentioned improved gaming experiences and better handling of resource-intensive tasks.

2. Value for Money: Several positive reviews highlighted the cost-effectiveness of the processors, particularly when compared to previous generations or competitors.
3. Cooling Efficiency: Improved thermal management was a recurring theme in positive reviews, with users appreciating the processors' ability to maintain performance under load.
4. Technological Advancements: Positive sentiment was often associated with the processors' support for new technologies and standards, enhancing overall system capabilities.

◇ **Negative reviews-**

1. Power Consumption: Some negative reviews focused on high power consumption, particularly in the context of the i9-14900K model.
2. Price Concerns: A subset of users felt that the price point was too high, especially when considering the incremental improvements over previous generations.
3. Compatibility Issues: There were mentions of compatibility problems with certain motherboards or other hardware components.
4. Thermal Management: While generally positive, some users reported challenges with managing heat output, especially under heavy loads.

- **Discussion**:
  - ◇ The sentiment analysis of Amazon reviews for Intel's 14th Gen processors provides valuable insights into consumer perceptions. The predominantly positive sentiment (69.7%) suggests that Intel has successfully met or exceeded customer expectations with this generation. The high frequency of terms like "performance," "gaming," and "cooler" in positive reviews indicates that Intel's focus on these areas has resonated well with consumers.
  - ◇ The analysis also revealed interesting temporal and regional patterns. For instance, there was a notable spike in positive sentiment in July 2024, possibly indicating a successful product launch or update. Additionally, the sentiment varied across different countries, with Spain showing the highest average sentiment (0.36) and France showing the lowest (-0.18).
  - ◇ The aspect-based sentiment analysis provided further nuanced insights. For example, performance-related aspects generally received positive sentiment, while price and power consumption were more contentious issues.

◇ The use of advanced NLP techniques like topic modeling and bigram analysis allowed for a deeper understanding of consumer concerns and praises. For instance, the frequent co-occurrence of terms like "power consumption" and "expensive" in negative reviews highlights areas where Intel might focus future improvements.

TABLE:

GENERATION	TOTAL NO OF REVIEWS COLLECTED	CORES	SENTIMENT DISTRIBUTION
11th Gen	703	i5-11400, i5-11400F, i5-11600K, i7-11700, i7-11700K, i9-11900K	Positive Reviews: 350 out of 703(49.8%), Negative reviews: 353 out of 703(50.2%)
12th Gen	1050	i9-12900K, i9-12900KS i5-12400F, i5-12400,i5-12600K i7-12700K, i7-12700F, i7-12700	Positive Reviews: 70.66%, Negative reviews: 10.95%, Neutral Reviews: 18.38%
13th Gen	979	i3-13100, i3-13100F, i5-13400, i5-13400F, i5-13500, i5-13600K, i5-13600KF, i7-13700, i7-13700F, i7-13700K, i7-13700KF, i9-13900, i9-13900F, i9-13900K, i9-13900KF	Positive Reviews: 65.56%, Negative reviews: 24.34%, Neutral Reviews: 10.1%
14th Gen	502	i3-14100F, i3-14100, i5-14400F, i5-14500, i7-14700F, i7-14700K, i9-14900KS, i9-14900K	Positive Reviews: 350 out of 502(69.7%), Negative reviews: 99 out of 502(19.7%), Neutral Reviews: 53 out of 502(10.6%)

Conclusion

Sentiment analysis for the 11th, 12th, 13th, and 14th generations of Intel processors has come up with diverse insights concerning perception and consumer satisfaction. With the methods, models, and feature extraction techniques in use, we attained a deep understanding of sentiment trends and their implications.

- 11th Gen: Intel processors are generally seen in a positive light. Models such as DistilBERT, Naive Bayes, SVM, and Random Forest were used in the sentiment analysis; DistilBERT performed particularly well at capturing contextual subtleties. The majority of opinions, according to the analysis, were favorable and emphasized increased performance, energy efficiency, and value for the money. Negative reviews, on the other hand, mostly addressed compatibility problems, cost issues, and a few early flaws. By addressing these issues, customer happiness may be increased even further. To preserve and raise customer satisfaction, future initiatives should focus on enhancing compatibility, resolving pricing issues, and guaranteeing consistent performance across all workloads.

- 12th Gen: Running the sentiment analysis on the fine-tuned RoBERTa model on the 12th generation processors returned a Sutton distribution, with most of them positive. The sentiment analysis revealed the major satisfaction in performance gaming and energy efficiency. But, product arrival, packaging, and ordinary quality of motherboards are usually shared issues in most negative reviews. Work on bettering the unboxing experience and quality would offset such negative sentiments.

- 13th Gen: Rule-based systems and machine learning algorithms—involving BERT and RoBERTa—have been used for the 13th generation processors. BERT turned out to be at the top with the most accuracy among all the models used for the analysis. The positive review includes performance improvement, energy efficiency, and technological advancement. The negative reviews focus on compatibility issues, pricing in comparison with rivals, and variability of performance. Focusing on the two issues of compatibility and pricing, these will help build better satisfaction among users.

- 14th Gen: In the case of 14th generation CPUs, sentiment analysis was done on a full scale with VADER, TextBlob, LDA, and TF-IDF for deep analysis. The positive sentiment drivers for this generation included increased performance, energy efficiency, and advanced features such as AI capabilities and integrated graphics. The negative feedback referred to compatibility, pricing, and performance under some workloads. These shall further improve customer satisfaction and loyalty.

Overall, Intel processors from all generations received very positive reviews. People have liked the performance boost, power efficiency, and technological advancement. On the negative note, recurring themes include incompatibility, high price, and bugs or inconsistent performance at an early stage. Intel can use these types of common issues to further improve and enhance consumer satisfaction, which will give them a competitive edge in the marketplace. These sentiment analysis methodologies provided insightful information concerning the perception of consumers and guided future product development and customer service initiatives.

Suggestions and Recommendations for Intel

The following suggestions and recommendations are to help alleviate common negative feedback and improve overall customer satisfaction:

- **Improve Compatibility and Integration:**
    1. More Extensive Compatibility Testing: Test for more motherboards and components for perfect integration.
    2. Quite Clearly Defined Compatibility Guidelines: Enough detail in compatibility guidelines and resources should be provided for the customer to avoid adverse cases.
    3. Firmware Updates: Regularly release firmware updates to resolve the issue related to compatibility effectively and with the best user experience.

- **Pricing Strategies Optimization:**
    1. Competitive Pricing: Adjust pricing strategies from time to time to maintain competitiveness vis-à-vis alternative products, considering the value proposition through performance and features.
    2. Tiered Pricing Models: Set up tiered pricing models aimed at different market segments. Appeal to price-sensitive consumers and offer premium versions to heavy users.

- **Improve Product Quality and Reliability:**
    1. Strengthen quality control measures with severe checks for low defects and high reliability in the line of products.
    2. Customer Feedback Integration: The prompt collection and integration of customer feedback into the product development process can help recognize common issues with the product and fix them.

- **Packaging and Delivery Experience Enhancements:**
    1. Protective Packaging: Spend more on safer and more protective packaging to prevent damage during transit. Thus, enhance the unboxing experience.
    2. Reliable Delivery Partners: Collaborate with reliable delivery services able to deliver products within time frames and ensure safety.

- **Address Initial Bugs and Performance Inconsistencies:**
    1. Thorough Testing: Include comprehensive pre-release testing procedures to detect and eliminate bugs before launch.

2. Prompt Bug Fixes: Make sure that there are teams for rapid responses to fix each and every bug or performance issue related to any report.
3. User Support: Provide robust and full-fledged user support channels for helping customers troubleshoot and resolve issues quickly.

- **Improved Customer Support and Communication:**
  1. 24/7 Support: Offer 24/7 customer support services to attend to any problem or issues that customers may encounter, as soon as possible.
  2. Ample communication: Keep the customers informed about any updates on products, patches, and fixes to reported problems.
  3. Educational Materials: Prepare educational resources such as tutorials and FAQs that can act as a guide for the user to harness the potential of the products.

- **Shift in Focus to More Sustainable Practices:**
  1. Eco-friendly Packaging: The companies should Rain eco-friendly materials for their packaging to gain attraction from green consumers.
  2. Energy Efficiency: More work on the power efficiency of processors, touting sustainability as part of marketing strategies.

- **Better Feature Set and Technological Innovations:**
  1. Innovative Features: Continue to introduce new and innovative features with an element of capturing changing consumer needs through AI capabilities and improved integrated graphics.
  2. Performance Optimization: Focus on workload performance optimization to ensure consistent and reliable performance across use cases.

Implementing these suggestions and recommendations, Intel shall have effectively addressed the negative feedback pointed in the sentiment analysis, improved customer satisfaction, and continued leading the processor market notwithstanding the highly competitive market.

Reference:

[1] Wikipedia. Link: https://en.wikipedia.org/wiki/Tiger_Lake
[2] Khan, F.H., Pasha, M.A., & Masud, S. (2021). "Advancements in Microprocessor Architecture for Ubiquitous AI—An Overview on History, Evolution, and Upcoming Challenges in AI Implementation," Micromachines, 12(6), 665. Link: https://www.mdpi.com/2072-666X/12/6/665
[3] Product Brief 11th Gen Intel® Core™ Desktop Processors. https://www.intel.com/content/dam/www/central-libraries/us/en/documents/11th-gen-product-brief-3-2-21.pdf
[4] 12th Gen Intel Core Processors on COM-HPC and COM Express Computer-on-Modules. Link: https://circuitcellar.com/research-design-hub/tech-trends/12th-gen-intel-core-processors-on-com-hpc-and-com-express-computer-on-modules/
[5] Intel Quietly Raises Prices for 12th-Gen Alder Lake CPUs, Now Cost More Than 13th-Gen. Link: https://www.tomshardware.com/news/intel-raises-pricing-for-12th-gen-alder-lake-processors-now-more-expensive-than-13th-gen
[6] Intel Announces Non-K 13th gen Core for Desktop: New 65W and 35W Processors. Link: https://www.anandtech.com/show/18702/intel-announces-non-k-13th-gen-core-for-desktop-new-65-w-and-35-w-processors
[7] Intel Announces non-K 14th Gen Core Desktop Processors: raptor Lake in 65 W to 35 W Flavors. Link: https://www.anandtech.com/show/21214/intel-announces-non-k-14th-gen-core-desktop-processors-raptor-lake-in-65-w-to-35-w-flavors