Intel Unnati

2024 - 2025

INTEL PRODUCT SENTIMENT ANALYSIS

Annual Report Insights

Prepared by:
TANISHA MUKHERJEE

From:
KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY

www.intel.in
support@unnatiindustrialtraining2024.com

ACKNOWLEDGEMENT

We extend our profound gratitude to all those who formally supported this project. First and foremost, we thank our institution, KIIT University, for coordinating this program and guiding us along the path of learning. Our deepest appreciation goes to our project supervisor and mentor, Mr. Debyut Hazra, for his invaluable guidance and expertise. We are indebted to Intel Unnati for providing us with opportunities and timely guidance throughout our journey. We also sincerely appreciate our department faculty for their professional input and the study participants for generously contributing their time and insights.

On a personal note, we would like to acknowledge the informal support that was equally vital to our success. Our families and friends offered unwavering encouragement and spread positivity throughout this challenging journey. Special thanks go to Shreyartha Roy for his late-night brainstorming sessions, assistance with problem-solving, and constant motivation. The countless cups of coffee prepared by my mother fueled our long work hours, and the understanding shown by our loved ones during our periods of absence was truly appreciated.

TABLE OF CONTENTS

Introduction

ABSTRACT

This study applies sentiment analysis techniques to user reviews of Intel products to gain insights into customer perceptions and satisfaction. Using a dataset of 3500 reviews collected from Amazon and Flipkart Customer review sections over 12 days , we employed NLTK, Flask Api and trained the model from EDA(Exploratory Data Analysis) to classify sentiment polarity and extract key topics. Our analysis revealed that positive and negative comments were there for analysis, we used unsupervised data in this case and amazingly we found that there were many mixed reviews of the processors but the graph showed the variations in sentiment analysis and we used a bar graph to showcase the behaviourial chart of the model. The results provide valuable feedback for product development and marketing strategies, highlighting areas of strength and potential improvement for Intel. This research demonstrates the effectiveness of sentiment analysis in understanding customer experiences and informing business decisions in the semiconductor industry.

BACKGROUND

Sentiment analysis, a subset of natural language processing, has become increasingly important in understanding customer opinions and experiences. In the competitive semiconductor industry, where Intel is a major player, customer feedback is crucial for product development and market positioning.

This project aims to leverage sentiment analysis techniques to extract valuable insights from user reviews of Intel products. By analyzing large volumes of unstructured text data, we can identify trends in customer satisfaction, highlight product strengths and weaknesses, and uncover specific areas for improvement.

Previous studies have shown the effectiveness of sentiment analysis in various industries, but its application to semiconductor products, particularly Intel's offerings, has been limited. This research fills that gap by focusing on Intel-specific reviews and adapting sentiment analysis techniques to the unique vocabulary and context of computer hardware discussions.

The findings from this study can inform Intel's strategic decisions, from product design to marketing strategies, ultimately enhancing customer satisfaction and maintaining competitiveness in the rapidly evolving tech market.

OBJECTIVE

- **Sentiment Classification**: Identifying whether the overall sentiment of a text is positive, negative, or neutral.
- **Sentiment Intensity Measurement**: Determining the strength or degree of the expressed sentiment.
- **Aspect-based Analysis**: Identifying sentiments toward specific aspects or features mentioned in the text.
- **Opinion Extraction**: Uncovering subjective information and attitudes expressed about particular topics or entities.

SCOPE

1. Data Sources:-Social media posts and comments,Customer reviews and feedback,Surveys and questionnaires,News articles and blogs,Emails and customer support interactions,Financial reports and earnings calls.
2. Industries and Sectors:-Retail and E-commerce,Finance and Investment,Healthcare,Hospitality and Tourism,Politics and Government,Entertainment and Media,Technology and Software
3. Linguistic Scope:-Single-language analysis,Multilingual sentiment analysis,Cross-lingual sentiment analysis.

4.Analytical Approaches:-Rule-based systems,Machine learning algorithms,Deep learning models,Hybrid approaches

5.Time Frame:-Real-time sentiment analysis,Historical entiment analysis,trend analysis over time.

6. Output Formats:-Numerical scores,Categorical labels,Visual representations (charts, graphs),Textual summaries

Integration:-Standalone sentiment analysis tools, Integration with larger analytics platforms,Incorporation into business intelligence systems

8.Ethical Considerations:-Privacy and data protection,Bias detection and mitigation,Transparency and explainability of algorithms

Literature Review

Theoritical Paradigm

1. Sentiment Polarity Theory:
   - The basic idea that text can be classified as positive, negative, or neutral
   - Forms the foundation of most sentiment analysis tasks
2. Bag-of-Words (BoW) Model:
   - Represents text as an unordered set of words
   - Simplifies text analysis by focusing on word occurrence
3. Tokenization:
   - The process of breaking text into individual words or tokens
   - A crucial first step in text processing
4. Stop Words:
   - Common words (like "the", "is", "at") that are often removed in text analysis
   - Helps focus on more meaningful words
5. Lexicon-based Approach:
   - Using pre-defined lists of words with associated sentiments
   - Provides a straightforward method for sentiment scoring
6. Rule-based Systems:
   - Simple if-then rules for classifying sentiment
   - Easy to understand and implement for beginners

7. Basic Text Classification:
   - The concept of categorizing text into predefined classes
   - Fundamental for supervised learning in sentiment analysis
8. Naive Bayes Classification:
   - A simple probabilistic classifier based on Bayes' theorem
   - Often used as an introductory machine learning method for text classification
9. Term Frequency (TF):
   - Measures how frequently a term occurs in a document
   - Basic concept for understanding word importance
10. Basic Data Preprocessing:
   - Techniques like lowercasing, removing punctuation, and handling special characters
   - Essential for cleaning and standardizing text data.

RESEARCHES

- Tech Product Sentiment Analysis: "Aspect-based sentiment analysis of online customer reviews using BERT" (2020) by Pei et al.t
- Social Media Sentiment for Tech Companies: "Twitter sentiment analysis: The good the bad and the omg!" (2011) by Kouloumpis et al.
- Product Feature Sentiment Analysis: "Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings" (2015) by Irsoy and Cardie.
- "Twitter Sentiment Analysis Using NLTK and Machine Learning Techniques" (2019) by Rahul Pandey et al.
- "Sentiment Analysis of Product Reviews Using NLTK" (2018) by Suresh Goutam.
- Real-time Sentiment Analysis for Product Launches: "Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle" (2012) by Wang et al.
- Aspect-Based Sentiment Analysis for Tech Products: "Aspect-based sentiment analysis on electronic products using BERT" (2020) by Xu et al.
- User Review Analysis for Electronics: "Mining and Summarizing Customer Reviews" (2004) by Hu and Liu.

Data Collection

DEFINITION

Data collection for sentiment analysis is a **structured process** that forms the basis of opinion mining in textual data. It involves gathering relevant information from diverse sources like **social media, e-commerce sites, customer feedback, forums, and news outlets**. Techniques such as web scraping, API integrations, surveys, and manual curation are used to compile user-generated content and formal texts expressing opinions on specific subjects or products. The process requires careful sampling, ethical considerations, and metadata capture. Collected data undergoes preprocessing and cleaning before being stored in analysis-friendly formats. This comprehensive effort creates a robust dataset for accurate sentiment analysis, providing insights into public opinion, customer satisfaction, and brand perception across various domains.

SOURCES

Data collection for sentiment analysis of Intel products focuses on three primary sources: **Amazon, Flipkart, and PC Benchmark**. Amazon and Flipkart, as major **e-commerce** platforms, provide a wealth of customer reviews, ratings, and discussions across various Intel products. PC Benchmark contributes specialized **technical evaluations and user comment**s, offering **in-depth** assessments of Intel hardware performance. These sources collectively offer a comprehensive view of consumer opinions and expert analyses, covering a broad

spectrum of Intel's product lineup and capturing both general user sentiment and technical reception. But we mainly focused on user reviews of intel core products.general user sentiment and technical reception. But we mainly focused on user reviews of intel core products.

METHODS

Natural Language Processing (NLP) tools:

- Text analysis software to process large volumes of data

Sentiment tracking over product lifecycle:

- Monitor changes in opinion from product launch through end-of-life

Predictive analytics:

- Use AI models to forecast sentiment trends based on historical data

Benchmarking data:

- Gather performance metrics and user reactions to them

STORAGE

Data is directly stored in **CSV** (Comma-Separated Values),**JSON** files after performing **Exploratory Data Analysis** (EDA). This approach is efficient for data manipulation and analysis. The process typically involves importing specialized data processing libraries such as **pandas** (often referred to as PANDA), **NumPy**, and utilizing various **Python** data structures like lists,**NLTK**, corpus.

Data Preprocessing

CLEANING TECHNIQUES

Modern data cleaning techniques leverage advanced AI language models like **ChatGPT, Monica, and Claude** to enhance efficiency and accuracy. These models excel at natural language processing tasks, making them valuable tools for text data cleaning. They can assist in identifying and correcting spelling and grammatical errors, standardizing text formats, and removing irrelevant content. ChatGPT, known for its versatility, can be used to paraphrase and normalize inconsistent text entries. Monica, with its focus on task completion, can help in categorizing and labeling data points. Claude, with its strong analytical capabilities, can aid in detecting anomalies and inconsistencies in large datasets. These AI assistants can also help in entity recognition, sentiment classification, and content summarization, streamlining the data cleaning process. By utilizing these advanced language models, data scientists can significantly reduce the time and effort required for data preparation, while also improving the overall quality and consistency of the dataset for subsequent analysis.

Methodology

- **Data Collection**: The data was gathered from multiple sources and consolidated into CSV files. This multi-source approach ensures a diverse dataset, potentially capturing a wide range of sentiments and contexts.
- **Data Cleaning**: Advanced AI language models, specifically ChatGPT and Monica, were employed to enhance the data cleaning process. These tools likely assisted in tasks such as text normalization, error correction, and standardization, significantly improving the quality and consistency of the dataset.
- **Model Selection and Training:** A combination of trained models was utilized, with Support Vector Machines (SVM) serving as the primary classifier. SVMs are well-suited for sentiment analysis due to their effectiveness in high-dimensional spaces and their ability to handle non-linear decision boundaries. Additionally, Long Short-Term Memory (LSTM) networks were incorporated to a lesser extent.
- **Dataset Size**: The models were trained on a substantial dataset of approximately 3,500 samples. This size strikes a balance between being large enough to capture diverse patterns and small enough for efficient processing.
- **Sentiment Analysis Process**: The trained models analyze the text data, classifying sentiments as positive or negative. The system then calculates the ratio of positive to negative sentiments, providing a quantitative measure of overall sentiment.
- **Output Visualization:** The results of the sentiment analysis are visualized using a bar graph. This graphical representation allows for quick and intuitive interpretation of the sentiment distribution.

Implementation

TOOLS

Jupyter Notebook was used as the primary development environment. Jupyter Notebook is an open-source web application that allows you to create and share documents containing live code, equations, visualizations, and narrative text. It's particularly well-suited for data analysis tasks due to its interactive nature and ability to combine code execution with rich text and visualizations.
The project required the importation of several key Python libraries:

1. **NumPy**: A fundamental package for scientific computing in Python. It provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays efficiently.
2. **Pandas**: A powerful data manipulation and analysis library. It offers data structures like DataFrames, which are particularly useful for handling structured data and performing operations such as merging, reshaping, and aggregating data.
3. **Matplotlib:** A comprehensive library for creating static, animated, and interactive visualizations in Python. It provides a MATLAB-like interface for generating plots and graphs.
4. **Seaborn**: A statistical data visualization library built on top of Matplotlib. It provides a high-level interface for drawing attractive statistical graphics and is particularly useful for creating complex visualizations minimal code.

5. **NLTK (Natural Language Toolkit)**: A leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning. Within NLTK, two specific modules were particularly important for this project:

a. **Stopwords**: This module provides lists of stop words (common words like "the", "a", "an", "in" that are often filtered out before processing natural language data) for various languages. Removing stop words is a common preprocessing step in many NLP tasks, including sentiment analysis.

b. **Corpus**: This module provides access to a variety of textcorpora, including categorized text collections, parsed sentences, and more. These corpora can be useful for training and testing NLP models.

The use of these libraries and modules in combination allows for efficient data manipulation, sophisticated natural language processing, and effective visualization of results. This comprehensive toolkit enables a robust approach to sentiment analysis, from data preprocessing and feature extraction to model training and result interpretation.

EVALUATION MATRIX

1. **Grammar and Interpretation Check:**

The system first evaluates the input text for grammatical accuracy and correct interpretation. Any content that fails this initial assessment is excluded from further analysis, ensuring the quality of the data being processed.

2. **Tokenization and Word Analysis:**

The text is then broken down into individual tokens or words. These tokens serve as the basis for sentiment classification.

3. **Sentiment Classification:**

The system analyzes the tokens to determine the overall sentiment:

- If more than half of the relevant tokens indicate positivity, the text is classified as positive.
- If the majority of tokens suggest negativity, the text is labeled as negative.
- In cases where there's no clear majority, the sentiment is categorized as mixed, indicating a balance of positive and negative elements.

4. **Multi-component Analysis and Visualization:**

When a CSV file containing multiple text components is uploaded, the system performs sentiment analysis on each component individually. The results are then aggregated and visualized in a bar graph format. This graphical representation provides a clear overview of sentiment distribution across all analyzed components..

MODEL IMPLICATION

Support Vector Machines (SVM) can be effectively integrated into an **NLTK-based** sentiment analysis project for beginners. The process typically begins with data preparation, where text data is collected and preprocessed using **NLTK's tools**. This involves tokenizing the text, **removing** stop words, and potentially applying **stemming** or lemmatization. The next crucial step is feature extraction, where the text is converted into a numerical format that SVM can process. A common approach is to use a **bag-of-word**s model, creating a vocabulary from the most frequent words in the dataset. Each text is then represented as a vector indicating the presence or absence of these words. NLTK provides functions to facilitate this process. Once the features are extracted, the data is split into **training and testing** sets. After training, the **model's performance** is evaluated using the test set, typically measuring metrics like accuracy, precision, and recall. Finally, the trained SVM model can be used to predict the sentiment of new, unseen text data. This approach provides a solid foundation for sentiment analysis, combining NLTK's powerful text processing capabilities with SVM's effective classification algorithm, making it accessible for beginners while still offering room for more advanced techniques as skills develop.Since we are working on **Unsupervised** ML algoritthm so **LSTM** is not used extensively.

Result and Discussion

RESULT WITH INSIGHT

Intel processors are generally high-quality and reliable. Specially **Intel core i7** of latest generation is exremly good and **85%** accepted whereas **Intel core i3** 10th gen processor is not that much compatible basically due to it's **quad core** technology for demanding workloads.Evo laptops offer a great balance of portability and performance. Sometimes it is observed that that laptops get hot under heavy loads and not a good choice **budget-conscious buyers**.There were more of mixed reviews where we got neutral analysis.But proper analysis of positive or negative depends on **customers needs and budget.**

USER SENTIMENTS

We have mixed reviews but depending on product and user's needs we got positive sentiments due to it's performance, reliability, high brand value and Evo laptops. Negative sentiments were mostly due it's requirement of heat management and frequent socket changes.

CONCLUSION

Intel: **A Throne Under Siege, But Not Toppled Yet**

For years, Intel has been the undisputed king of the CPU game. Gamers coveted their raw processing power, and professionals relied on their unwavering stability. But the tide is turning. AMD's fiery Ryzen processors offer similar or better performance at times, often at a more tempting price point. Digging deeper, sentiment analysis can unlock a treasure trove of insights. By analyzing the "whys" behind user opinions, Intel can identify features that resonate and areas for improvement. Imagine - products that not only meet user needs but anticipate them!

FUTURE RECOMMENDATION

Innovation is the game at Intel and they are concentrating on some aspects crucial to their coming products. They include:

Process Technology: Intel is developing new manufacturing processes like Intel 20A and 18A which will deliver significant performance and efficiency.

Artificial Intelligence: Intel is integrating neural processing units (NPU) into their processors to boost AI capabilities. This will allow for faster AI processing and new user experiences on PCs.

Data Center Products: Through Sierrra Forest and Granite Rapids, they are making a switch to Xeon's next generation of processors, which would do a lot in improving data centers' performances as well as energy efficiency.

Graphics Technology: With Arrow Lake the company approaching desktop processors will be bringing its Arc graphics architecture. Thus, this will provide better graphics performance for desktop.

Foundry Powerhouse: Intel is heavily investing in becoming a major foundry, manufacturing chips designed by other companies. This could diversify their revenue stream and solidify their position in the chipmaking ecosystem.