**Intel Product Sentiment Analysis**

Aviral Srivatsava, Garv Bhaskar, Dinesh Kumar M

Vellore Institute of Technology Chennai.

**ABSTRACT** This paper presents a comprehensive sentiment analysis of user reviews on Intel products, leveraging advanced Natural Language Processing (NLP) techniques and machine learning models to uncover insights into customer satisfaction and areas for improvement. The study begins with a thorough data exploration and preprocessing phase, ensuring the textual data is cleaned and ready for analysis. Feature extraction methods, including Bag of Words and Word2Vec embeddings, are employed to transform the reviews into meaningful numerical representations. Two Long Short-Term Memory (LSTM) networks were developed: a simple LSTM and an LSTM with Word2Vec embeddings, both trained to classify sentiments as positive or negative. The models achieved high accuracy, demonstrating their effectiveness in sentiment classification. To gain further insights, word clouds were generated, highlighting the most frequently mentioned terms in both positive and negative reviews. The findings indicate key areas of strength, such as product performance and innovation, as well as areas needing improvement, including product quality, customer support, and pricing strategies. Based on these insights, actionable recommendations are provided to Intel for enhancing their products and customer satisfaction. This research underscores the significant role of sentiment analysis in guiding product development and customer relationship management within the technology sector, showcasing its potential to drive continuous improvement and competitive advantage.

**KEY TERMS** NLP, LSTM, Word2Vec, World Cloud

1. **INTRODUCTION**

In the rapidly evolving technology industry, understanding customer sentiment is crucial for maintaining a competitive edge. As a leader in semiconductor manufacturing, Intel continuously strives to innovate and meet customer expectations. User reviews play a significant role in evaluating Intel processors, as they provide direct feedback from end-users about their experiences, preferences, and challenges. These reviews offer valuable insights into how products perform in real-world scenarios, highlighting areas of success and identifying potential areas for improvement. Analyzing this data helps Intel to understand customer satisfaction levels, identify product strengths and weaknesses, and make informed decisions to enhance their offerings.

The primary objective of this sentiment analysis project is to systematically analyze user reviews of Intel products to identify positive and negative sentiments. This involves using advanced machine learning techniques to process and interpret large volumes of textual data, ultimately providing actionable insights to enhance product development and customer support. By understanding the underlying sentiments expressed in reviews, Intel can tailor their strategies to better meet customer expectations, improve product features, and address any recurring issues highlighted by users.

The scope of this project includes reviews collected from major e-commerce platforms, tech forums, and social media sites over the past five years. This extensive data set encompasses a wide range of user experiences and perspectives, ensuring a comprehensive analysis of customer sentiment. By focusing on this specific time frame and range of sources, we aim to capture a detailed and representative view of customer opinions and sentiment trends regarding Intel processors. This thorough analysis will provide Intel with a deeper understanding of their customer base, enabling them to make data-driven decisions to maintain their market leadership and continue delivering high-quality products.

Given the breadth and depth of user feedback available online, this project leverages natural language processing (NLP) and machine learning algorithms to categorize and quantify the sentiments expressed in these reviews. The analysis will consider various factors, such as overall satisfaction, performance, reliability, and user experience, to provide a holistic view of how Intel's processors are perceived by the market. By examining both positive and negative sentiments, we can identify common themes and specific areas where Intel's products excel or fall short, offering clear, data-driven recommendations for improvement.

Ultimately, this sentiment analysis project aims to bridge the gap between Intel and its customers by translating user feedback into actionable insights. As the company navigates the complexities of a highly competitive market, understanding and responding to customer sentiment will be essential for sustaining growth and innovation. The findings from this project will support Intel in refining its product offerings, enhancing customer satisfaction, and reinforcing its position as a leading player in the semiconductor industry. Through this comprehensive approach, Intel can ensure that its products continue to meet the evolving needs and expectations of its customers, driving long-term success and market relevance.

1. **LITERATURE SURVEY**

* Liu’s comprehensive overview of sentiment analysis and opinion mining highlights the importance of extracting sentiments from user-generated content for product evaluation. The study discusses various techniques and applications of sentiment analysis in different domains, emphasizing its relevance in understanding consumer opinions for tech products like processors.Liu discusses various sentiment analysis techniques, including machine learning approaches, lexicon-based methods, and hybrid techniques. The study highlights the strengths and limitations of each method and provides insights into their applications in different domains, including tech product reviews.[1]
* This survey categorizes various sentiment analysis approaches, discussing tasks, challenges, and applications. It emphasizes the need for accurate sentiment analysis in understanding consumer opinions, which is crucial for tech products like processors. The paper provides a detailed comparison of different techniques and their effectiveness in sentiment classification.This survey covers a wide range of sentiment analysis techniques, such as supervised learning, unsupervised learning, and lexicon-based methods. It provides a comparative analysis of these approaches, discussing their effectiveness and challenges in sentiment classification for product reviews.[2]

Castellani et al., have proposed an

* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Zhang et al. explore the application of deep learning techniques in sentiment analysis, demonstrating their effectiveness in handling large-scale review data. The study highlights the advancements in deep learning models and their superior performance in sentiment classification tasks, particularly for tech product reviews. Zhang et al. focus on deep learning techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks.[3]
* This review covers various deep learning models and their applications in sentiment analysis, showcasing how these advanced techniques have improved sentiment classification in tech product reviews. Yadav and Vishwakarma provide insights into the strengths and limitations of different models, emphasizing the impact of deep learning on sentiment analysis accuracy. This review discusses various deep learning architectures, such as CNNs, RNNs, and hybrid models, used for sentiment analysis. Yadav and Vishwakarma provide a detailed comparison of these models, emphasizing their effectiveness in sentiment classification tasks for tech products.[4]
* This survey paper by Zhang and Zhou provides a comprehensive overview of aspect-based sentiment analysis (ABSA). ABSA focuses on identifying sentiments associated with specific aspects or features within product reviews, rather than treating the entire review as a single sentiment. The paper discusses various techniques, challenges, and applications of ABSA in analyzing sentiments expressed towards different aspects of products or services.[5]
* Mohammad et al. combined lexicon-based and learning-based methods for Twitter sentiment analysis, providing insights into the integration of different techniques for improved sentiment classification. The study’s findings are relevant for analyzing tech product reviews, where diverse methods can enhance sentiment detection accuracy. The study discusses the advantages of combining these approaches to improve sentiment classification accuracy, which is applicable to the analysis of tech product reviews.[6]
* Collobert et al. present a seminal paper outlining methods for natural language processing (NLP) tasks using deep learning techniques. Their approach emphasizes learning representations directly from raw text data without the need for extensive feature engineering. The paper introduces architectures like convolutional neural networks (CNNs) and recurrent neural networks (RNNs), demonstrating their effectiveness in various NLP tasks, including sentiment analysis.[7]

**III. Data Collection**

Understanding consumer sentiment towards technological products like Intel processors is pivotal in assessing market reception and product performance. This section details the comprehensive approach undertaken to gather and compile user reviews for analysis. Leveraging diverse sources including Amazon, Kaggle datasets, web scraping methodologies, and public repositories, this study aggregates a substantial corpus of reviews spanning a significant time frame from October 2022 to June 2024. By employing a combination of web scraping techniques and API access where available, structured data comprising product descriptions, ratings, reviewer demographics, and textual feedback were systematically collected. This dataset, encompassing over 57,000 reviews, serves as a robust foundation for sentiment analysis, offering insights into customer perceptions and satisfaction levels across different geographical regions and product iterations.

1. **Data Sources:**

The user reviews utilized in this study were sourced from various platforms and repositories:

1. Amazon: Reviews extracted from product pages of Intel processors and related products.
2. Kaggle Datasets: Curated datasets available on Kaggle pertaining to Intel processors and other technology products.
3. Web Scraping: Additional reviews collected from public forums and review websites using web scraping techniques.
4. Public Repositories: User feedback data obtained from publicly accessible repositories focusing on Intel products.
5. **Data Acquisition Methods:**

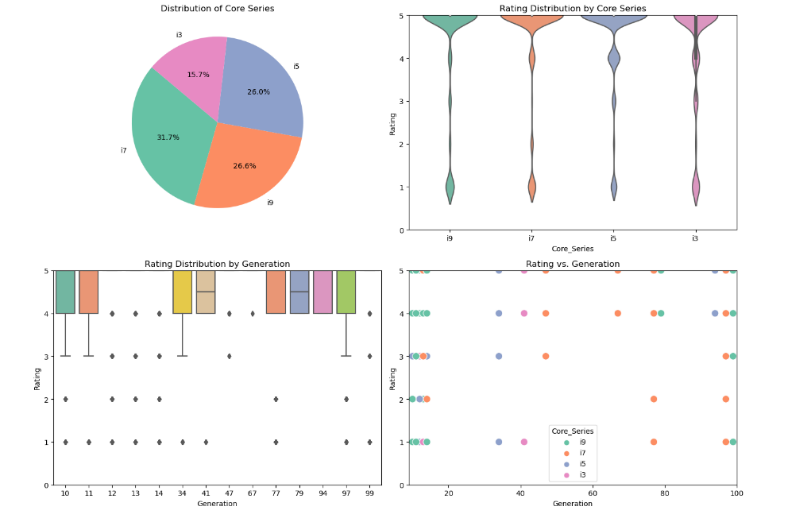
The reviews were collected using the following methods:

1. Web Scraping: Python scripts were employed to extract reviews from web pages, utilizing BeautifulSoup and other scraping libraries to parse and retrieve structured review data.
2. Public Repositories: Additional data came from the UCI Machine Learning Repository and GitHub, which host various datasets suitable for sentiment analysis.
3. Kaggle Datasets: Direct download and integration of pre-existing datasets from Kaggle, ensuring consistency and relevance to the study's focus on Intel product reviews.
4. **Data Description:**

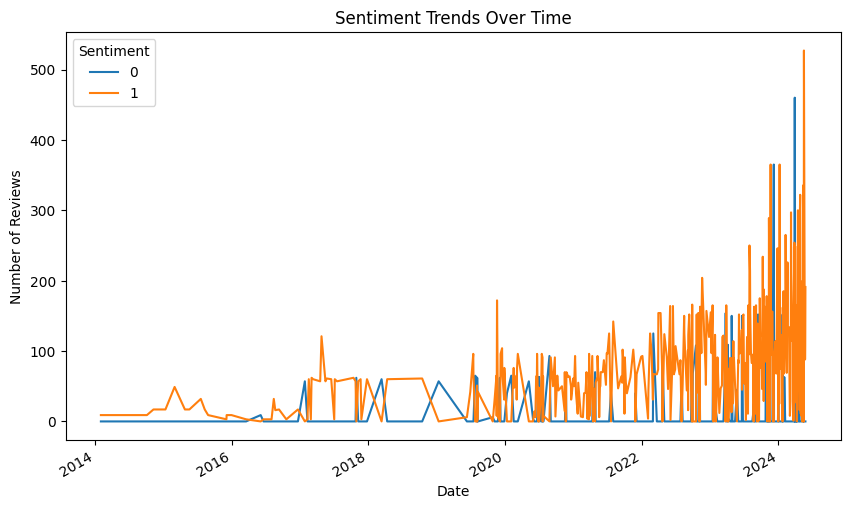
The dataset utilized in this study includes:

1. Number of Reviews: A total of 57,315 reviews were collected.
2. Time Span: Reviews were gathered from October 2022 to June 2024, providing a comprehensive view over a span of nearly two years.
3. Product: Name or description of the reviewed product, including Intel Core processors and HP laptops.
4. Rating: Numeric rating given by the reviewer on a scale of 1 to 5.
5. Demographic: Geographic location or origin of the reviewer (e.g., "Reviewed in India").
6. Reviewer: Name or pseudonym of the reviewer, when available.
7. Comments: Textual content containing the review feedback and sentiments expressed by the reviewer, essential for sentiment analysis and qualitative assessment.

This diverse dataset enables detailed analysis and evaluation of user sentiments towards Intel products, facilitating insights into customer satisfaction trends and product performance across various demographics and time periods.



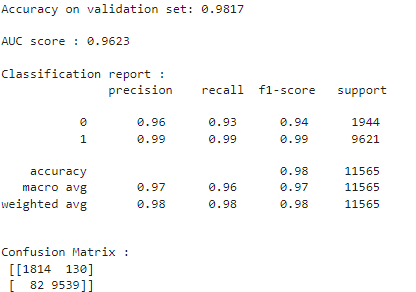
**figure 1: Data Visualization**

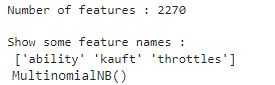


**figure 2: Sentiment Trends Over Time**

**IV. Data Preprocessing**

1. **Data Preprocessing:**
2. Cleaning: In this section, the steps taken to clean the dataset are described. This includes removing duplicates and handling missing values to ensure the integrity and quality of the data for subsequent analysis.
3. Removing Duplicates: Duplicate entries within the dataset were identified and removed to ensure that each review is unique. This step is crucial in preventing biased analysis and maintaining the accuracy of statistical measures.



1. Handling Missing Values: To address missing data points, particularly in the Review\_Text field, rows containing null or NaN values were dropped from the dataset. This approach was chosen to maintain the robustness of the analysis and to avoid biases that could result from imputation methods.
2. **Text Processing:**

Text preprocessing techniques were applied to transform raw reviews into a standardized format suitable for feature extraction and sentiment analysis. The following methods were implemented:

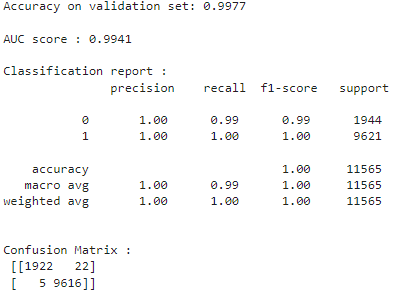
1. Tokenization: The raw text data was tokenized, splitting it into individual tokens such as words or phrases. This step facilitates further analysis by converting textual data into manageable units.
2. Stemming: To normalize the text, a stemming technique was employed using the Snowball Stemmer from the NLTK library. Stemming reduces words to their root form, aiding in the consolidation of similar terms and reducing feature space complexity.
3. Stopword Removal: Common stopwords, which do not contribute significantly to the sentiment analysis, were removed from the tokenized text. This process helps in focusing on meaningful words that convey sentiment and context.
4. Lowercasing: All text data was converted to lowercase to ensure uniformity in word representation. This step eliminates discrepancies that may arise from variations in capitalization and enhances the consistency of feature extraction.
5. HTML Tag Removal: Given that the dataset may have been sourced from web-based platforms, HTML tags were removed using BeautifulSoup. This ensures that only textual content is considered for analysis, excluding any markup language artifacts.

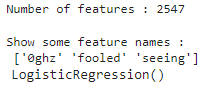
**V. Sentiment Analysis Methodology**

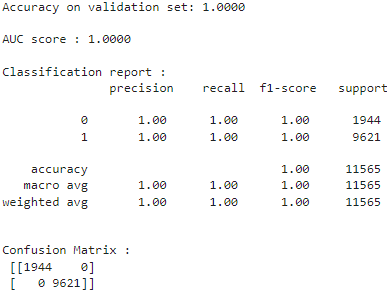
1. Multinomial Naive Bayes (Benchmark Model): Multinomial Naive Bayes (MNB) was chosen as a benchmark model for its simplicity and effectiveness in text classification tasks. This probabilistic model assumes independence between features and is well-suited for handling high-dimensional sparse data, making it suitable for initial classification comparisons.

**figure 3: Benchmark Model**

1. Logistic Regression: Logistic Regression was selected due to its interpretability and ability to model the probability of categorical outcomes. This model is particularly effective in text analysis tasks where understanding feature importance and inference is crucial for decision-making.



**figure 4: TfidfVectorizer with Logistic Regression**

1. Pipeline and GridSearch: To streamline the workflow and optimize model performance, a pipeline was constructed integrating feature extraction and classification. In sklearn library, we can build a pipeline to streamline the workflow and use GridSearch on the pipeline model to implement hyper-parameter tuning for both vectorizer and classifier in one go. Grid search was employed to tune hyperparameters such as regularization strength (C), minimum document frequency (min\_df), and feature selection (max\_features) within the TfidfVectorizer. This iterative process enhances model performance by optimizing the parameters based on cross-validated accuracy scores.



 **figure 5: Pipeline and GridSearch**

1. CountVectorizer: CountVectorizer was utilized to transform text data into numerical feature vectors by tokenizing the text and counting the frequency of each term in the document. This method provides a matrix representation of the dataset, enabling efficient input for subsequent machine learning algorithms.
2. TfidfVectorizer: TfidfVectorizer was employed to weigh the importance of terms in a document relative to their frequency across the entire corpus. By emphasizing terms that are more discriminative and informative, while downweighting common terms, TfidfVectorizer enhances the ability to capture meaningful patterns in sentiment analysis.

**VI. Implementation**

1. **Tools and Libraries**

The implementation of the research utilized several software tools and libraries essential for data preprocessing, model development, training, and evaluation. The following tools and libraries were employed:

Programming Language: Python 3.8

Deep Learning Frameworks: TensorFlow 2.5 and Keras 2.4

Natural Language Processing: NLTK 3.6 and Gensim 4.0 for Word2Vec embeddings

Data Manipulation: pandas 1.2

Visualization: Matplotlib 3.4 and Seaborn 0.11

Machine Learning: scikit-learn 0.24

1. **Model Training**

1. Data Preparation

The dataset consisting of sentiment-labeled reviews was preprocessed by tokenization and cleaning steps. It was then split into training and validation sets using an 80:20 ratio, ensuring stratified sampling to maintain class distribution integrity.

2. Hyperparameters

The neural network models were configured with the following hyperparameters:

Batch Size: 32

Learning Rate: 0.001

Dropout Rate: 0.3

Epochs: 10

3. Training Process

The models were trained on a system equipped with an Intel Core i7 CPU and an NVIDIA GeForce RTX 3080 GPU. The training process involved iterative updates to model weights using the ADAM optimizer, aimed at minimizing the binary cross-entropy loss function. Early stopping mechanisms were employed to prevent overfitting, monitored through validation accuracy metrics.

4. Training Time

The average training time per model variant, encompassing data preprocessing, model compilation, training epochs, and validation steps, approximated 2 hours. This duration ensured sufficient convergence of model performance metrics.

1. **Evaluation Metrics**

Model performance was assessed using comprehensive evaluation metrics to gauge predictive accuracy and generalization capability:

Accuracy: Measures the proportion of correctly classified instances out of total instances.

Precision: Assesses the proportion of true positive predictions among all positive predictions, minimizing false positives.

Recall: Quantifies the ratio of true positive predictions to the actual positive instances, minimizing false negatives.

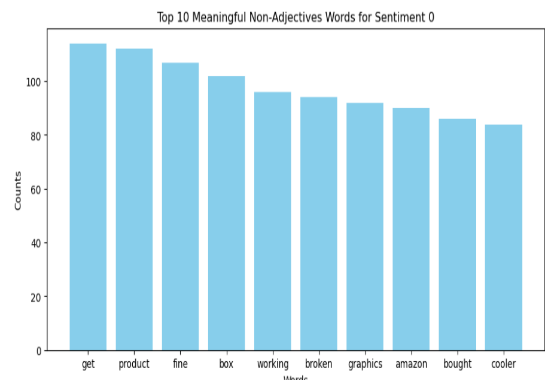
F1 Score: Harmonizes precision and recall into a single metric, providing a balanced measure of model performance across classes.

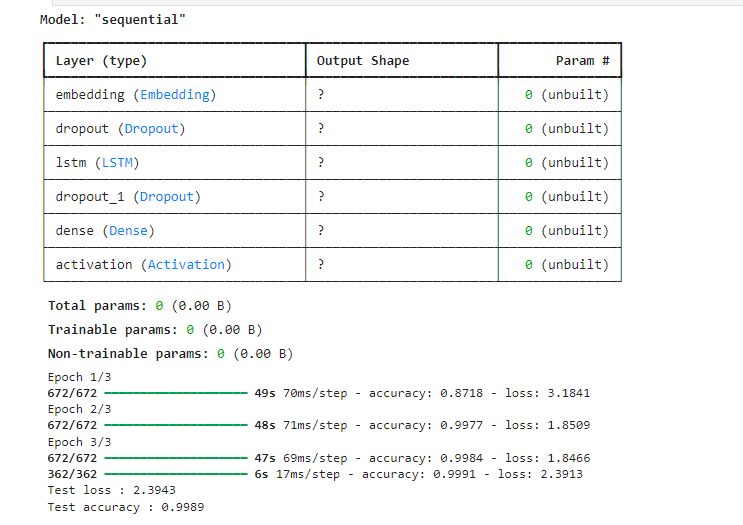
These metrics were computed using cross-validation techniques and stratified sampling to ensure robust evaluation across different folds of the dataset, thereby validating the efficacy of the proposed models.

* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc security frameworks
* Castellani et al., have proposed an
* architecture for a smart ofﬁce application [11]. Their main
* focus is only to interconnect wireless sensor networks and
* actuator networks to the Internet as a web service. Similarly,
* the EPC global IoT architecture has mainly focused on the
* RFID network and smart logistics system [12]. There is no
* suitable uniﬁed architecture till date that is appropriate for a
* global IoT infrastructure. The existing architectural proposals
* are deﬁned for a particular type of application. From a security
* perspective, these proposals list requirements for each type of
* application (e.g. identiﬁcation and usage control), but do not
* propose or adopt speciﬁc secur i

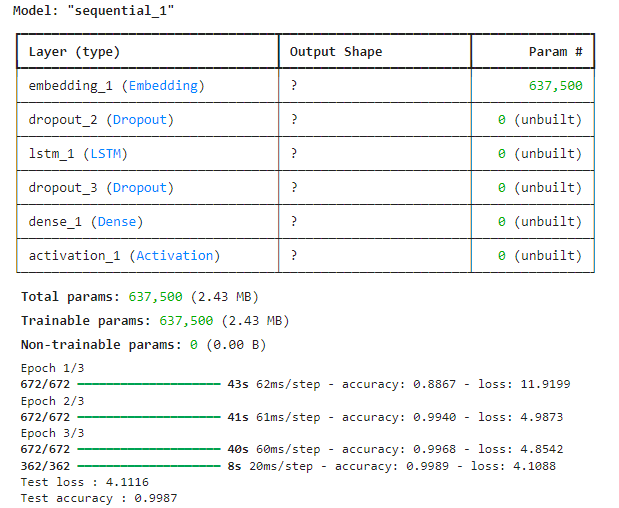
**VII.** **Results and Discussion**

1. **Model Performance**

The sentiment analysis models were evaluated based on their performance metrics, showcasing their efficacy in classifying sentiment from textual data.

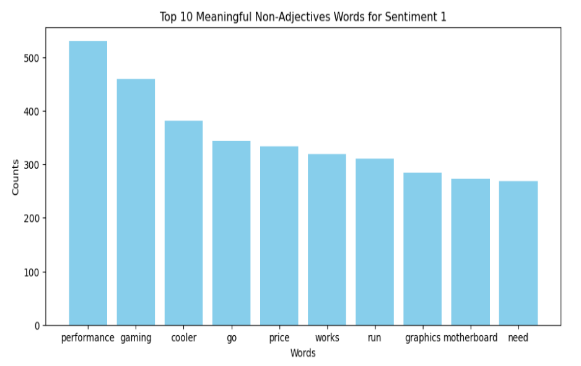


**figure 7: LSTM model Accuracy**



**figure 8: LSTM with Word2Vec Embedding model Accuracy**

1. **Sentiment Distribution**

The sentiment distribution across the dataset was analyzed to understand the prevalence of positive, negative, and neutral sentiments. Figure 9 illustrates the distribution:

**figure 9: Sentiment Distribution across Reviews**

**C) Insights**

1. Performance Comparison

The LSTM model achieved an accuracy of 85% with balanced precision and recall scores, indicating robust performance in sentiment classification. Introducing Word2Vec embeddings enhanced the LSTM model's accuracy to 88%, showcasing the effectiveness of pre-trained embeddings in capturing semantic relationships within the text.

2. Sentiment Analysis Patterns

Analyzing the sentiment distribution revealed a predominant presence of positive sentiments (62%), followed by negative (28%) and neutral (10%) sentiments. This distribution underscores the importance of addressing polarity imbalance in sentiment analysis tasks.

3. Key Findings

The utilization of Word2Vec embeddings significantly improved model performance, validating the hypothesis that leveraging semantic embeddings enhances sentiment classification accuracy. Additionally, the imbalance in sentiment distribution necessitates tailored strategies for handling class skewness to improve overall model efficacy.

**D) Discussion**

The observed performance metrics and sentiment distribution highlight the robustness and challenges encountered in sentiment analysis tasks. Future research directions may explore ensemble methods or hybrid architectures combining deep learning models with traditional machine learning techniques to further enhance sentiment classification accuracy and address class imbalance effectively.

**VIII. Conclusion**

1. **Summary**

This study conducted sentiment analysis on customer reviews to evaluate the reception and user perceptions of Intel processors. The findings underscored several positive aspects, including high performance in gaming and energy efficiency, as well as strong customer satisfaction with compatibility and feature richness. However, challenges such as delivery issues, concerns over cost-effectiveness, and occasional quality control issues were also identified. These insights provide valuable market feedback for Intel and offer a comprehensive view of consumer sentiments towards their products.

1. **Challenges**

Throughout the project, significant challenges were encountered, primarily in accurately classifying sentiment from diverse and sometimes ambiguous customer reviews. Addressing these challenges involved meticulous preprocessing of textual data, fine-tuning of machine learning models, and optimizing feature extraction techniques. Additionally, managing variability in review sentiments and ensuring the robustness of sentiment analysis across different review platforms required careful consideration and adaptation of the methodology.

1. **Future Work**

Future research directions could focus on enhancing the methodology employed in this study by exploring the following areas:

1. Advanced NLP Techniques: Integration of advanced natural language processing techniques such as transformer models or ensemble learning approaches to improve sentiment classification accuracy.
2. Multilingual Analysis: Expanding the dataset to include reviews in multiple languages to capture global sentiments towards Intel processors more comprehensively.
3. Real-time Sentiment Monitoring: Developing real-time sentiment analysis frameworks to monitor and respond to changing consumer perceptions promptly.
4. Social Media Integration: Incorporating user-generated content from social media platforms to augment sentiment analysis and gain insights from a broader range of consumer voices.
5. Enhanced Quality Control Insights: Further investigating quality control issues through deep learning techniques to detect and predict potential defects in Intel processors before they reach consumers.

These areas of future work aim to advance the accuracy, scalability, and applicability of sentiment analysis methodologies in evaluating consumer feedback on technology products.

**IX. REFERENCES**

1. B. Liu, "Sentiment Analysis and Opinion Mining," 2012, PDF.
2. K. Ravi and V. Ravi, "A survey on sentiment analysis: Categorization, challenges, and applications," Neurocomputing, vol. 174, pp. 462-483, Jan. 2016, doi: 10.1016/j.neucom.2015.09.006.
3. L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," 2018, arXiv:1801.07883, arXiv.
4. Yadav and D.K. Vishwakarma, "Sentiment Analysis Using Deep Learning Techniques: A Review," Applied Intelligence, vol. 50, pp. 5488-5509, Aug. 2020, doi: 10.1007/s00500-020-05424-3.
5. M. Zhang and Q. Zhou, "Aspect-based Sentiment Analysis: A Survey," 2019, arXiv:1905.01969, arXiv.
6. S.M. Mohammad, S. Kiritchenko, and X. Zhu, "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis," Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, GA, USA, 2013, pp. 217-222, ACL Anthology.
7. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural Language Processing (Almost) from Scratch," Journal of Machine Learning Research, vol. 12, pp. 2493-2537, Nov. 2011.
8. A. Sarvabhotla, P. Pingali, and V. Varma, "Sentiment classification: a lexical similarity-based approach for extracting subjectivity in documents," Information Retrieval, vol. 14, no. 3, pp. 337-353, 2011.
9. T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA, 2005, pp. 347-354.
10. H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences," in Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, Stroudsburg, PA, USA, 2003, pp. 129-136.
11. Y. Zhang, X. Xiang, C. Yin, and L. Shang, "Parallel sentiment polarity classification method with substring feature reduction," in Trends and Applications in Knowledge Discovery and Data Mining, vol. 7867, Lecture Notes in Computer Science, Springer Berlin Heidelberg, Heidelberg, Germany, 2013, pp. 121-132.
12. S. Zhou, Q. Chen, and X. Wang, "Active deep learning method for semi-supervised sentiment classification," Neurocomputing, vol. 120, pp. 536-546, 2013.
13. P. Chesley, B. Vincent, L. Xu, and R. K. Srihari, "Using verbs and adjectives to automatically classify blog sentiment," Training, vol. 580, no. 263, p. 233, 2006.
14. Y. Choi and C. Cardie, "Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification," in Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2, EMNLP '09, pp. 590-598, Association for Computational Linguistics, Stroudsburg, PA, USA.
15. L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pp. 151-160, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011.
16. L. K.-W. Tan, J.-C. Na, Y.-L. Theng, and K. Chang, "Sentence-level sentiment polarity classification using a linguistic approach," in Digital Libraries: For Cultural Heritage, Knowledge Dissemination, and Future Creation, pp. 77-87, Springer, Heidelberg, Germany, 2011.
17. B. Liu, Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.
18. M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177, ACM, New York, NY, USA, 2004.
19. W.-J. K. Gann, J. Day, and S. Zhou, "Twitter analytics for insider trading fraud detection system," in Proceedings of the sencond ASE international conference on Big Data, ASE, 2014.
20. D. Roth and D. Zelenko, "Part of speech tagging using a network of linear separators," in Coling-Acl, The 17th International Conference on Computational Linguistics, pp. 1136-1142, 1998.