IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

Intelligent Sentiment Analysis of Customer Reviews Using NLP

Prof. S V Raut[1], Mr. Shiva Wanjalkar[2], Mr. Gaurav Janbandhu[3], Mr. Sayed Aadib[4], Ms. Divya Belekar[5], Ms.Gauri Dayrekar[6]
Assistant Professor, Department of Computer Science & Engineering[1]
Undergraduate Students, Department of Computer Science & Engineering[2,3,4,5,6]
Dr. Rajendra Gode Institute of Technology and Research, Amravati (MH), India

Abstract: *The current explosion of e-commerce activity has created a massive data bottleneck where thousands of daily customer reviews are generated, making manual oversight physically impossible for businesses. This Study addresses the challenge by developing an automated sentiment analysis pipeline specifically designed to process and categorize unstructured consumer feedback into positive, negative, or neutral sentiments. Utilizing a substantial corpus of over 200,000 product reviews, the system employs a rigorous Natural Language Processing (NLP) workflow that prioritizes lemmatization over basic stemming to preserve the semantic integrity of the text. By transforming the cleaned data through TF-IDF vectorization incorporating both unigrams and bigrams the model is trained to capture critical contextual nuances, such as negation, which simpler frequency-based models often misinterpret. For the classification stage, the Multinomial Naive Bayes algorithm was selected due to its high computational efficiency and proven performance with high-dimensional textual datasets. To bridge the gap between theoretical modeling and real-world application, the final system was integrated into a Flask-based web interface, providing an accessible platform for real-time sentiment prediction. Our experimental results, based on a 20% unseen test set, confirm that the model effectively distinguishes between polar sentiments while highlighting the inherent linguistic difficulty of classifying neutral feedback. Ultimately, this project provides a scalable, low-cost framework for small-to-medium enterprises to automate their customer feedback loops without requiring expensive infrastructure.*

Keywords: Sentiment Analysis

I. INTRODUCTION

The modern retail landscape has shifted almost entirely toward a feedback-driven economy, where e-commerce platforms serve as the primary marketplace for global consumers. Unlike traditional brick-and-mortar shopping, online consumers rely heavily on a collective digital conscience the product review section. These thousands of daily reviews are not just text; they represent a goldmine of data regarding product defects, usability success, and overall brand reputation. However, the sheer volume of this unstructured textual data presents a massive hurdle. For any growing business, it is physically and economically impossible to hire a human team to manually read, categorize, and act upon every single piece of feedback generated every minute. This creates a critical demand for automated Natural Language Processing (NLP) systems that can "read" and understand human emotion at scale. Sentiment Analysis, often referred to as Opinion Mining, provides the technical bridge to solve this problem by automatically classifying text into positive, negative, or neutral categories. While recent advancements in Deep Learning and Large Language Models (LLMs) offer high accuracy, they often come with massive computational costs and high latency, making them impractical for small-to-medium enterprises (SMEs) that need fast, low-cost solutions. This research focuses on building a more resource-efficient yet robust pipeline using the Multinomial Naive Bayes algorithm. We chose this approach because of its proven speed and effectiveness in handling high-dimensional text data, especially when paired with TF-IDF (Term Frequency-Inverse Document Frequency)

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, April 2026 Impact Factor: 8.2

vectorization. The objective of this project is to move beyond theoretical accuracy and build a deployable sentiment classification framework

II. LITERATURE REVIEW

The field of Sentiment Analysis, or Opinion Mining, has transitioned significantly from basic keyword matching to sophisticated probabilistic modeling. Early research in this domain was primarily lexicon-based, relying on pre-defined dictionaries of positive and negative words. While foundational, these methods as noted in early studies by Pang and Lee (2008) often failed to account for context, sarcasm, or linguistic shifts. A word like "long" might be positive for "battery life" but negative for "wait time," a distinction that static lexicons struggled to maintain. This gap led to the adoption of Machine Learning (ML) as a more dynamic solution for understanding consumer sentiment. The evolution of sentiment analysis reflects the broader progression of Natural Language Processing (NLP) from simple keyword matching to complex probabilistic modeling. This section categorizes previous research into four critical domains:

Early attempts at opinion mining relied heavily on predefined dictionaries of positive and negative words. While foundational, researchers like Pang and Lee (2008) demonstrated that these lexicon-based methods were notoriously rigid. They struggled to account for context where a word like "unpredictable" could be a compliment for a movie plot but a criticism for a laptop's battery. This limitation shifted the academic focus toward Supervised Machine Learning, where models learn emotional polarity directly from labeled datasets. Landmark studies by Pang, Lee, and Vaithyanathan (2002) proved that algorithms like Naive Bayes and SVM could outperform human-coded rules by identifying subtle statistical patterns in consumer language. In the specific context of high-volume text classification, computational speed is as vital as accuracy. McCallum and Nigam (1998) conducted a pivotal comparison of Naive Bayes variants, concluding that the Multinomial Naive Bayes (MNB) model is exceptionally suited for document-level analysis. Unlike the Bernoulli model which only checks for word presence, MNB accounts for word frequency, making it ideal for long product reviews. Recent literature continues to favor MNB for real-time applications because it requires significantly less training data and hardware power compared to modern "Black Box" deep learning models, while maintaining a Raw consumer feedback is often "noisy," filled with slang, typos, and grammatical irregularities that can confuse a classifier. Research in the field of text mining emphasizes that preprocessing is not just a cleanup step but a core requirement for model performance. While early systems used "Stemming" (which crudely chops word endings), modern NLP practitioners advocate for Lemmatization. By using morphological analysis to return words to their dictionary root (e.g., "better" to "good"), lemmatization preserves the semantic intent of the customer. This ensures that the model recognizes different tenses and pluralities of the same sentiment as a single, unified feature. Machine learning algorithms cannot process raw text; it must be converted into a mathematical "Vector Space." The TF-IDF (Term Frequency-Inverse Document Frequency) technique, as explored by Ramos (2003), remains the industry standard for this transformation. It effectively penalizes high-frequency "filler" words (like "the" or "product") and boosts the weight of sentiment-heavy terms (like "defective" or "excellent"). Furthermore, contemporary research highlights the importance of N-Grams specifically Bigrams to capture negation. Analyzing words in pairs (e.g., "not happy") prevents the model from misclassifying negative reviews that contain positive individual words, a common error in basic unigram models.

III. RESEARCH METHODOLOGY

The proposed system is built on a structured pipeline designed to convert raw, unstructured customer feedback into a refined numerical format for probabilistic classification. The methodology is divided into four critical phases:

The foundation of this research is a large-scale corpus consisting of over 200,000 product reviews sourced from an e-commerce platform. Unlike smaller datasets, this volume allows the model to encounter a vast diversity of

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, April 2026 Impact Factor: 8.2

consumer vocabulary, including informal slang and product-specific jargon. Each entry in the raw CSV format includes metadata such as Product ID, star ratings (1–5), and the primary review text. To prepare this for supervised learning, we mapped these numerical ratings into three categorical labels: Positive (4–5 stars), Neutral (3 stars), and Negative (1–2 stars). This labeling serves as the "ground truth" for training the classifier and evaluating its predictive accuracy.

3.1 Multistage Text Preprocessing and Normalization

Raw text is inherently "noisy" and contains many elements that do not contribute to sentiment. Our preprocessing pipeline follows a rigorous sequence to ensure high data quality:

➢ Case Normalization: All text is converted to lowercase to ensure that "excellent" and "excellent" are treated as the same feature.

➢ Noise Removal: We utilized Regular Expressions (Regex) to strip HTML tags, special characters, and numerical digits that carry no emotional weight.

➢ Tokenization & Stop-Word Filtration: Sentences were broken into individual tokens, and common "filler" words (e.g., "is," "the," "at") were removed using the NLTK library. This significantly reduces the dimensionality of the data.

➢ Morphological Lemmatization: Instead of crude stemming, which often results in non-dictionary words, we applied lemmatization to return words to their dictionary root (e.g., "better" to "good"). This ensures semantic consistency across the entire dataset.

Since machine learning algorithms require numerical input, we transformed the processed tokens into a Vector Space Model using TF-IDF (Term Frequency-Inverse Document Frequency). Unlike simple word counts, TF-IDF assigns a higher weight to words that are rare across the document but frequent in a specific review (like "faulty" or "superb"), effectively highlighting sentiment-bearing terms. To further enhance the model's contextual awareness, we incorporated Bigrams (two-word sequences). This prevents the common "Negation Error" where a model might misclassify "not good" as positive simply because it sees the word "good" in isolation. For the classification engine, we implemented the Multinomial Naive Bayes (MNB) algorithm. This probabilistic model is chosen for its exceptional speed in handling large-scale sparse matrices produced by TF-IDF. The dataset was split using an 80/20 ratio for training and testing to ensure unbiased validation. After the model reached its optimal performance metrics, it was serialized and integrated into a Flask web framework. This deployment allows for a functional front-end where users can input any review string and receive a real-time sentiment prediction, bridging the gap between a static experiment and a deployable software tool.

IV. RESULTS AND DISCUSSION

The performance of the developed sentiment analysis system was evaluated using a 20% hold-out test set, comprising approximately 40,000 unseen reviews. This section provides a critical analysis of the statistical outcomes, the model's strengths, and the inherent challenges in classifying natural language. The Multinomial Naive Bayes (MNB) classifier, when paired with TF-IDF vectorization, demonstrated high computational efficiency with a training time of under 15 seconds for 160,000 records. While the overall accuracy was high, we looked beyond simple metrics to analyze the Confusion Matrix. The model showed exceptional Precision for the Positive class, correctly identifying reviews that used distinct, high-polarity words like "superb," "authentic," and "reliable." However, we observed a slight dip in Recall for the Neutral class. This is primarily because neutral feedback often lacks strong emotional keywords, making it mathematically difficult for a probabilistic model to distinguish it from subtle negative sentiment. A significant finding in our research was the drastic reduction in "False Positives" after incorporating Bigrams into the feature extraction process. In initial trials using only Unigrams, phrases such as "not good" were frequently misclassified as positive because the model only registered the high statistical weight of the word "good." By expanding the vocabulary to include word pairs, the system

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, April 2026 Impact Factor: 8.2

successfully captured the negation context. This improvement was particularly vital for the Negative sentiment class, where customers often use sarcastic or negated positive terms to express dissatisfaction. Upon manual inspection of the misclassified test cases, we identified a recurring pattern of "Sentiment Polarity Ambiguity." This occurred mostly in 3-star reviews where a customer might praise one aspect of a product while criticizing another for example: "The camera quality is amazing, but the battery life is disappointing." Since our current model treats the entire review as a single bag-of-words, it occasionally struggles with such multi-faceted feedback. This highlights the limitation of document-level classification and suggests that future iterations could benefit from aspect-level sentiment mining. To validate the model's practical utility, we conducted real-time testing through the developed Flask web application. The system achieved an average inference latency of less than 85ms per prediction, proving its suitability for live e-commerce environments. Even when subjected to long, multi-paragraph reviews, the preprocessing pipeline effectively condensed the input into the top 5,000 most informative features without significant loss of context. This confirms that a lightweight MNB model is a viable, low-cost alternative to heavy deep-learning architectures for standard business monitoring.

V. CONCLUSION

The primary objective of this research was to bridge the gap between high-volume e-commerce data and actionable business insights by developing a scalable, automated sentiment analysis framework. By utilizing a substantial dataset of over 200,000 product reviews, we have demonstrated that while modern deep learning models often dominate academic discussions, the Multinomial Naive Bayes algorithm remains a highly potent and resource-efficient tool for real-world production environments. Our methodology successfully transformed noisy, unstructured consumer feedback into a refined numerical format through a rigorous pipeline of lemmatization and TF-IDF vectorization, specifically incorporating bigrams to tackle the persistent challenge of linguistic negation. The integration of this trained model into a Flask-based web application further validates the system's practical utility, achieving a remarkably low inference latency of under 85ms, which is critical for live monitoring in small-to-medium enterprises. While the model showed exceptional precision in identifying polarized sentiments, the experimental results also highlighted the inherent difficulty in classifying "neutral" feedback and multi-faceted reviews where conflicting opinions are present. This research concludes that a well-optimized probabilistic pipeline offers a viable, low-cost alternative to expensive GPU-dependent architectures, providing businesses with a reliable mechanism to monitor customer satisfaction and product performance in real-time. Moving forward, the framework established here serves as a foundation for more granular, aspect-based sentiment mining that could eventually decipher the most complex nuances of human consumer behavior.

VI. ACKNOWLEDGMENT

The successful completion of this research work on the Intelligent Sentiment Analysis of Customer Reviews Using NLP would not have been possible without the support and guidance of several individuals and organizations.

We would like to express our sincere gratitude to our project guide and mentors for their valuable suggestions, continuous encouragement, and insightful feedback throughout the development of this project. Their expertise and guidance played a crucial role in shaping this research work.

We are also thankful to our institution and faculty members for providing the necessary resources, infrastructure, and academic support required to carry out this study effectively.

Furthermore, we would like to acknowledge the contributions of the developer communities and open-source platforms that provided essential tools, libraries, and documentation, which greatly facilitated the implementation of this system.

Finally, we extend our heartfelt thanks to our family and friends for their constant support, motivation, and encouragement throughout the research process.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, April 2026
Impact Factor: 8.2

REFERENCES

[1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.
[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques," Proceedings of EMNLP, pp. 79–86, 2002.
[3] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, 2012.
[4] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proceedings of ACL, 2002.
[5] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," Proceedings of ACM SIGKDD, 2004.
[6] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," Stanford University Technical Report, 2009.
[7] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global Vectors for Word Representation," Proceedings of EMNLP, 2014.
[8] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," Proceedings of ICLR, 2013.
[9] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), 2017.
[10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedi