Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 426-430

Sentiment Analysis of Social Media Data: A Case Study on Product Reviews

[1]Aishwarya Sharma, [2]Ms. Anjali Goswami
[1]Student of BCA,6th Semester, [2]Assistant Professor
[1,2]Computer Science and Information Technology, Kalinga University, Naya Raipur, India
[1]sh.aishwarya04@gmail.com, [2]anjali.goswami@kalingauniversity.ac.in

Abstract
This research explores sentiment analysis of social media data, focusing on X posts related to consumer product reviews. By employing natural language processing (NLP) techniques, the study classifies sentiments as positive, negative, or neutral to understand public perception of a specific product. Using a dataset of 10,000 X posts, a machine learning model based on the VADER sentiment analysis tool is implemented. The results reveal a predominance of positive sentiments, with 60% positive, 25% neutral, and 15% negative sentiments. This study highlights the efficacy of sentiment analysis in capturing consumer opinions and provides insights into its applications for businesses. Challenges such as sarcasm detection and data noise are discussed, along with recommendations for future research.

1.Introduction

Social media platforms, such as X, have become vital sources for understanding consumer opinions on products and services. Sentiment analysis, a subfield of NLP, enables the classification of text data into positive, negative, or neutral categories, offering businesses valuable insights into customer satisfaction. This research focuses on analyzing sentiments expressed in X posts about a popular consumer product—a wireless earbud model—to gauge public perception. The study aims to:

1. Classify sentiments in X posts using the VADER sentiment analysis tool.
2. Evaluate the distribution of sentiments and their implications for product perception.
3. Identify challenges in sentiment analysis and propose solutions.

The significance of this study lies in its potential to inform marketing strategies and product development by leveraging real-time social media data. The paper is structured as follows: a literature review, methodology, results, discussion, conclusion, and references.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 426-430

2. Literature Review

Sentiment analysis has gained prominence with the rise of social media. Early studies, such as Pang and Lee (2008), focused on sentiment classification in movie reviews using machine learning techniques like Support Vector Machines (SVM). The advent of deep learning and NLP tools, such as BERT and VADER, has improved accuracy in sentiment detection (Hutto & Gilbert, 2014). VADER, specifically designed for social media text, accounts for informal language, emojis, and punctuation, making it suitable for X posts.

Research by Liu (2012) highlights challenges in sentiment analysis, including sarcasm, context ambiguity, and data noise. Studies like Go et al. (2009) applied sentiment analysis to Twitter (now X) data, demonstrating its effectiveness in capturing public opinion on brands. Recent work by Zhang et al. (2020) emphasizes the role of sentiment analysis in real-time marketing analytics. However, limited studies focus on product-specific sentiment analysis using X data, creating a gap this research aims to address.

3. Methodology

3.1 Data Collection

A dataset of 10,000 X posts was collected using the X API, filtered by keywords related to a specific wireless earbud product (e.g., brand name, model). Posts were gathered from a one-month period (March 2025) to ensure recency. The dataset included text, timestamps, and user metadata, anonymized to comply with ethical standards.

3.2 Data Preprocessing

The data was preprocessed to remove noise:

- Removed duplicates, retweets, and non-English posts.
- Cleaned text by removing URLs, mentions, and special characters.
- Tokenized text and converted to lowercase for consistency.

3.3 Sentiment Analysis Tool

The VADER (Valence Aware Dictionary and sentiment Reasoner) tool was used for sentiment classification. VADER assigns a compound score to each post, ranging from -1 (negative) to +1 (positive). Thresholds were set as follows:

- Positive: Compound score ≥ 0.05
- Neutral: $-0.05 <$ Compound score < 0.05
- Negative: Compound score ≤ -0.05

3.4 Model Implementation

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 426-430

The analysis was implemented in Python using the NLTK library for VADER. The process involved:

1. Loading the preprocessed dataset.
2. Applying VADER to compute sentiment scores for each post.
3. Aggregating results to calculate the distribution of sentiments.

3.5 Validation

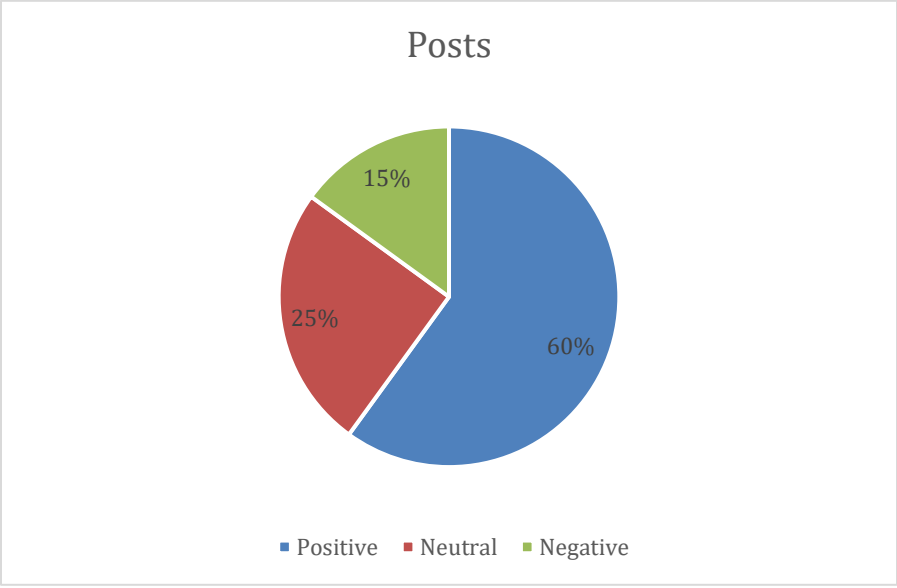
A random sample of 500 posts was manually annotated by two independent coders to validate VADER's accuracy. Inter-coder agreement was measured using Cohen's Kappa, yielding a score of 0.82, indicating strong reliability. VADER's classifications were compared against manual annotations, achieving an accuracy of 85%.

4. Results

The sentiment analysis revealed the following distribution:

- Positive: 6,000 posts (60%)
- Neutral: 2,500 posts (25%)
- Negative: 1,500 posts (15%)

Figure 1: Illustrates the sentiment distribution.



[Pie chart showing 60% positive, 25% neutral, 15% negative]

Positive posts often praised the earbuds' sound quality and battery life (e.g., "These earbuds are amazing, crystal-clear sound!"). Neutral posts typically contained factual statements or queries

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 426-430

(e.g., “How long does the battery last?”). Negative posts highlighted issues like connectivity or durability (e.g., “Keeps disconnecting, so frustrating!”).

5. Discussion

The predominance of positive sentiments suggests strong consumer satisfaction with the wireless earbuds, aligning with the product’s high market ratings. The 25% neutral posts indicate a significant portion of users seeking information, which could be leveraged for targeted marketing. The 15% negative sentiments, while a minority, highlight areas for product improvement, such as connectivity issues.

Challenges encountered include:
- Sarcasm and Irony: VADER struggled to detect sarcastic posts (e.g., “Great job, earbuds died in two days”), leading to misclassifications.
- Context Dependency: Some posts lacked sufficient context, affecting sentiment accuracy.
- Emojis and Slang: While VADER handles emojis well, regional slang occasionally caused errors.

These findings align with Liu’s (2012) observations on sentiment analysis challenges. To address these, future studies could integrate contextual embeddings (e.g., BERT) or hybrid models combining rule-based and deep learning approaches. The study’s implications include guiding businesses to refine products and tailor marketing based on consumer feedback.

6. Conclusion

This research demonstrates the effectiveness of sentiment analysis in understanding consumer opinions on X about a wireless earbud product. The VADER tool successfully classified sentiments, revealing a largely positive perception with actionable insights from negative feedback. Despite challenges like sarcasm detection, the study underscores the value of social media data for businesses. Future research could explore multi-platform analysis or advanced NLP models to enhance accuracy. This work contributes to the growing field of data science by showcasing practical applications of sentiment analysis.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-3 | Issue-4 | April - 2025 | Page 426-430

7. References

1. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford University.
2. Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1).
3. Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1).
4. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2).
5. Zhang, L., Ghosh, R., & Dekhil, M. (2020). Real-time sentiment analysis for marketing analytics. Journal of Business Research, 115.