

A Hybrid Sentiment Analysis Framework for Footwear Product Reviews Using Lexicon-Based and Machine Learning Techniques

Ajike Uchenna O., Awodele Oludele, Ajayi Oluwabukola F.

Date of Submission: 08-04-2026

Date of Acceptance: 21-04-2026

Abstract

Accurate sentiment analysis of online customer reviews is essential for data-driven decision-making in the footwear industry. Traditional lexicon-based approaches often struggle with contextual and domain-specific sentiment, while machine learning models require extensive labeled data. This paper proposes a hybrid sentiment analysis framework that integrates lexicon-based polarity scores with supervised machine learning techniques to improve sentiment classification accuracy for footwear product reviews. Using a publicly available dataset, lexicon-based sentiment scores from TextBlob and VADER were combined with TF-IDF features and used to train Logistic Regression, SVM, and AdaBoost classifiers. Experimental results demonstrate that the hybrid model significantly outperforms standalone approaches, achieving an accuracy of 90% and improved precision, recall, and F1-score. The proposed framework offers a robust and scalable solution for sentiment analysis in domain-specific e-commerce applications.

Keywords

Hybrid Sentiment Analysis, Footwear Reviews, Machine Learning, Lexicon-Based Methods, AdaBoost

I. Introduction

The rapid expansion of e-commerce has transformed how consumers evaluate and purchase products. In the footwear industry, customer reviews significantly influence purchasing behavior, brand perception, and product development strategies (Liu, 2012; Pang & Lee, 2008). These reviews contain valuable sentiment information that reflects customer satisfaction, preferences, and concerns. However, the growing volume of online reviews makes manual analysis impractical, necessitating automated sentiment analysis techniques.

Sentiment analysis, a subfield of natural language processing (NLP), focuses on identifying and classifying opinions expressed in textual data (Pang & Lee, 2008). Existing approaches broadly fall into lexicon-based and machine learning-based categories. While lexicon-based methods rely on

predefined sentiment dictionaries, machine learning approaches learn sentiment patterns from labeled data. Understanding customer sentiment is crucial for product development, quality control, and marketing strategies. However, accurately capturing sentiment from unstructured text remains a challenge due to linguistic ambiguity and domain-specific expressions. This study introduces a hybrid sentiment analysis framework designed to address these challenges.

II. Related Work

Sentiment analysis has been extensively studied within the field of natural language processing, with early research focusing on machine learning-based classification of opinions in text (Pang et al., 2002; Pang & Lee, 2008). Lexicon-based approaches, which rely on predefined sentiment dictionaries, have also been widely adopted due to their simplicity and interpretability (Turney, 2002; Liu, 2012). Tools such as VADER and TextBlob have demonstrated effectiveness in general-purpose sentiment analysis, particularly for short and informal text (Hutto & Gilbert, 2014; Loria, 2020).

More recent studies have explored hybrid sentiment analysis models that combine lexicon-based features with supervised machine learning techniques in order to overcome the limitations of individual approaches. Empirical findings indicate that hybrid models often achieve superior performance, especially in domain-specific applications where contextual meaning and specialized vocabulary are prevalent (Aggarwal & Zhai, 2012). This paper builds upon existing research by proposing a domain-specific hybrid framework tailored to footwear product reviews, addressing challenges related to contextual sentiment interpretation and classification accuracy.

III. Proposed Hybrid Framework

3.1 System Architecture

The proposed framework integrates lexicon-based polarity scores with machine learning features.

TextBlob and VADER sentiment scores are extracted and appended to TF-IDF feature vectors.

3.2 Model Training

The enhanced feature set is used to train Logistic Regression, SVM, and AdaBoost models. Hyperparameter tuning was performed to optimize classification performance.

4. Experimental Setup

The dataset was split into training and testing sets using an 80:20 ratio. Performance was evaluated using accuracy, precision, recall, and F1-score.

5. Results and Analysis

The hybrid AdaBoost model achieved the highest performance, with an accuracy of 90%, precision of 0.89, recall of 0.88, and F1-score of 0.88. These results demonstrate the effectiveness of combining lexicon-based sentiment information with machine learning techniques.

6. Implications for the Footwear Industry

The proposed hybrid framework enables footwear companies to extract actionable insights from customer reviews, leading to improved product design, enhanced customer satisfaction, and increased revenue.

7. Conclusion and Future Work

This study presents a hybrid sentiment analysis framework that outperforms traditional approaches in analyzing footwear product reviews. Future work will explore deep learning techniques and real-time sentiment analysis applications.

unsupervised classification of reviews. *ACL*, 417–424.

- [8]. Aggarwal, C. C., & Zhai, C. (2012). *Mining Text Data*. Springer.

References

- [1]. Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- [2]. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proceedings of EMNLP*, 79–86.
- [3]. Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of ICWSM*, 216–225.
- [4]. Loria, S. (2020). TextBlob: Simplified text processing. *Journal of Open Source Software*, 5(46), 2194.
- [5]. Joachims, T. (1998). Text categorization with Support Vector Machines. *European Conference on Machine Learning*, 137–142.
- [6]. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139.
- [7]. Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to