International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03770
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Hybrid Machine Learning Approaches for Enhanced Sentiment Analysis

Dr. P. Sumalatha
Assistant Professor
Dept. of Artificial Intelligence and Data Science
Central University of Andhra Pradesh
Ananthapuramu, India
sumalatha.psl@gmail.com

Jogu Umamahesh
Dept. of Artificial Intelligence and Data Science
Central University of Andhra Pradesh
Ananthapuramu, India
joguumamahesh@gmail.com

***Abstract*—This paper presents a hybrid sentiment analysis system that integrates lexicon-based, traditional machine learning, and deep learning techniques to classify textual data into positive, negative, or neutral sentiments. The system leverages Python-based libraries such as Scikit-learn, NLTK, and Transformers to pre-process text, extract features, and apply models including Support Vector Machine (SVM), Bidirectional Encoder Representations from Transformers (BERT), Linear Regression, and Valence Aware Dictionary and sEntiment Reasoner (VADER). The framework processes customer reviews from e-commerce platforms (Amazon, Flipkart) and social media (Instagram) using web scraping and provides actionable insights through sentiment summarization and visualizations (bar and pie charts). Experimental results demonstrate BERT's superior performance with 92.3% accuracy, followed by SVM (85.6%), Linear Regression (81.2%), and VADER (76.8%). The system addresses challenges like sarcasm, class imbalance, and scalability, offering a scalable, user-friendly solution for real-world applications in e-commerce, social media analytics, and brand reputation management.**

***Index Terms*—Sentiment Analysis, Machine Learning, Deep Learning, Natural Language Processing, BERT, SVM, VADER, Text Classification, Data Visualization.**

I. INTRODUCTION

Sentiment analysis (SA), a key area in natural language processing (NLP), aims to automate the classification of textual data into sentiments such as positive, negative, or neutral [1]. The proliferation of user-generated content on platforms like social media and e-commerce websites has necessitated scalable, automated systems for analyzing large-scale text data [2]. Existing systems often struggle with sarcasm detection, multilingual text, and domain-specific nuances, limiting their effectiveness [3]. This paper proposes a hybrid sentiment analysis system combining traditional machine learning (ML), deep learning (DL), and lexicon-based approaches to address these challenges. The system employs models like SVM, BERT, Linear Regression, and VADER, integrated with advanced NLP techniques and visualizations to provide actionable insights.

The rapid growth of digital communication platforms has exponentially increased the volume of textual data available for analysis, making sentiment analysis (SA) an indispensable tool for businesses, policymakers, and researchers [4]. By extracting subjective information from text, SA enables organizations to gauge public opinion, monitor brand reputation, and derive insights from customer feedback in real-time [5]. However, the complexity of human language, characterized by contextual subtleties, cultural variations, and non-literal expressions such as irony or sarcasm, poses significant challenges to achieving high accuracy in sentiment classification [6]. Traditional approaches, while effective in controlled environments, often fail to generalize across diverse datasets or adapt to the dynamic nature of online discourse [7]. This necessitates the development of more robust and versatile systems capable of addressing these limitations.

The proposed hybrid sentiment analysis system seeks to overcome these challenges by leveraging the strengths of multiple methodologies. Traditional machine learning models, such as Support Vector Machines (SVM) and Linear Regression, provide interpretable and computationally efficient solutions for structured datasets [8]. Meanwhile, deep learning models like BERT (Bidirectional Encoder Representations from Transformers) excel in capturing contextual relationships and semantic nuances in unstructured text, making them particularly suited for complex datasets [9]. Additionally, lexicon-based approaches, such as VADER (Valence Aware Dictionary and sEntiment Reasoner), offer a lightweight yet effective method for sentiment scoring, especially in scenarios where computational resources are limited or rapid processing is required [10]. By integrating these approaches, the proposed system aims to achieve a balance between accuracy, scalability, and interpretability.

A key innovation of this system lies in its ability to handle domain-specific nuances and multilingual text, which are critical for real-world applications. For instance, sentiment expressed in product reviews may differ significantly from that in political discourse or social media conversations, requiring tailored preprocessing and feature extraction techniques [11]. Furthermore, the system incorporates advanced NLP techniques, such as named entity recognition (NER) and part-of-speech (POS) tagging, to enhance its understanding of context and improve sarcasm detection. These techniques enable the system to disambiguate ambiguous terms and identify sentiment shifts within a single text [12]. To further aid decision-making, the system integrates visualization tools to present sentiment trends and patterns in an intuitive manner, empowering stakeholders to derive actionable insights from complex datasets.

The motivation for this hybrid approach stems from the limitations of relying solely on one methodology. For example, lexicon-based methods, while fast and intuitive, often struggle with context-dependent sentiments and fail to capture the subtleties of informal language prevalent on social media [13]. Conversely, deep learning models, despite their superior performance in contextual analysis, require substantial computational resources and large labeled datasets, which may not always be available [14]. By combining these methods with traditional ML, the proposed system aims to mitigate their individual shortcomings while maximizing their collective strengths. This paper outlines the system's architecture, evaluates its performance across diverse datasets, and demonstrates its applicability in real-world scenarios such as e-commerce, social media monitoring, and public opinion analysis.

Future sections of this paper are organized as follows: Section II reviews related work in sentiment analysis, highlighting gaps in existing approaches. Section III details the proposed hybrid system's methodology, including data preprocessing, model integration, and visualization techniques. Section IV presents experimental results, comparing the system's performance against baseline models on benchmark datasets. Section V discusses practical applications and limitations, while Section VI concludes with future research directions. Through this comprehensive approach, the proposed system aims to advance the field of sentiment analysis by providing a scalable, accurate, and adaptable solution for analyzing the ever-growing landscape of textual data.

II. Related Work

Sentiment analysis has evolved from lexicon-based and rule-based methods to data-driven ML and DL approaches [4]. Mamani-Coaquira and Villanueva [5] highlight the shift towards neural network-based models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers, which outperform traditional methods in capturing contextual relationships. Obiedat et al. [6] address class imbalance using hybrid SVM with Particle Swarm Optimization (PSO) and oversampling techniques like SMOTE, achieving improved performance on Arabic restaurant reviews. However, gaps remain in aspect-level analysis, multilingual support, and real-time scalability [5], [6]. This study builds on these findings by integrating multiple models and addressing domain-specific challenges through a hybrid framework.

The advent of machine learning (ML) techniques marked a significant shift toward data-driven sentiment analysis. Algorithms like Support Vector Machines (SVM), Naive Bayes, and Logistic Regression have been widely adopted for their robustness and ability to handle structured features, such as term frequency-inverse document frequency (TF-IDF) vectors [10]. Studies like those by Pang et al. [11] demonstrated the effectiveness of ML models in classifying movie reviews, achieving competitive accuracy with relatively low computational overhead. However, these models often require extensive feature engineering and struggle with capturing long-range dependencies in text, particularly in datasets with diverse linguistic patterns or informal language [12]. To address these limitations, researchers have explored hybrid approaches, such as combining SVM with optimization techniques like Particle Swarm Optimization (PSO) or Genetic Algorithms, as seen in Obiedat et al. [6], which improved classification performance on imbalanced datasets.

The introduction of deep learning (DL) models has further revolutionized sentiment analysis by enabling the automatic extraction of complex features from raw text. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been effective in modeling sequential data, capturing temporal dependencies in text [13]. More recently, Transformer-based models like BERT and its variants (e.g., RoBERTa, DistilBERT) have set new benchmarks by leveraging bidirectional context and attention mechanisms to understand nuanced semantic relationships [14]. For example, Devlin et al. [15] demonstrated BERT's superior performance on sentiment classification tasks across multiple languages, attributed to its ability to model context at a granular level. Despite their success, DL models are computationally intensive and require large annotated datasets, which may not be feasible for low-resource languages or specialized domains [16]. Additionally, their "black-box" nature can hinder interpretability, a critical factor in applications requiring transparency, such as legal or financial sentiment analysis.

Recent research has also focused on addressing specific challenges in sentiment analysis, such as sarcasm detection, aspect-based sentiment analysis (ABSA), and multilingual processing. Sarcasm, for instance, poses a significant challenge due to its reliance on tone and context, often leading to misinterpretations by lexicon-based and traditional ML models [17]. Advanced techniques, such as incorporating emoji analysis or contextual embeddings from models like ELMo or BERT, have shown promise in improving sarcasm detection [18]. Similarly, ABSA, which involves identifying sentiments toward specific aspects of an entity (e.g., service quality in restaurant reviews), has gained traction. Works like those by Sun et al. [19] combine LSTM with attention mechanisms to achieve fine-grained sentiment classification, though scalability remains a concern for real-time applications. Multilingual sentiment analysis, meanwhile, faces challenges due to linguistic diversity and the scarcity of labeled datasets in low-resource languages. Cross-lingual transfer learning, using models like mBERT or XLM-R, has been explored to address this, but performance varies across language pairs [20].

Despite these advancements, several gaps persist in the literature. Real-time scalability remains a critical issue, particularly for applications involving high-velocity data streams, such as social media or live customer feedback [21]. Additionally, most existing systems are designed for specific domains or languages, limiting their generalizability. For instance, models trained on English movie reviews may underperform on multilingual social media datasets or domain-specific texts like medical reviews [22]. Furthermore, class imbalance, as highlighted by Obiedat et al. [6], continues to challenge model performance, particularly in datasets with skewed sentiment distributions. Techniques like SMOTE or cost-sensitive learning have been proposed, but their effectiveness varies across contexts [23].

This study builds on these prior works by proposing a hybrid sentiment analysis framework that integrates lexicon-based, ML, and DL approaches to address the identified gaps. By combining the interpretability of VADER, the robustness of SVM, and the contextual awareness of BERT, the proposed system aims to provide a versatile solution capable of handling sarcasm, multilingual text, and domain-specific nuances. Additionally, the framework incorporates advanced preprocessing techniques, such as named entity recognition (NER) and dependency parsing, to enhance context understanding and improve aspect-level analysis. The system also emphasizes scalability through optimized model ensembles and real-time visualization tools, making it suitable for diverse applications, from e-commerce to public opinion monitoring. The following sections detail the system's architecture, evaluation methodology, and comparative performance against state-of-the-art approaches.

III. Methodology

The proposed system follows a quantitative experimental design to develop and evaluate a hybrid sentiment analysis framework. The methodology includes data collection, preprocessing, feature extraction, model development, and performance evaluation.

A. *Data Collection*

The system uses two data sources:

- **Static Dataset**: A curated dataset of 1,000 labeled Amazon reviews (balanced between positive and negative sentiments) for training and validation [7].
- **Dynamic Scraping**: Real-time extraction of reviews from Amazon, Flipkart, and Instagram using Selenium WebDriver and BeautifulSoup [8].

B. *Text Preprocessing*

Raw text is processed through a pipeline involving:

- HTML tag removal, special character handling, and case normalization.
- Tokenization, stopword removal, lemmatization, emoji handling, and contraction expansion [9].
- For BERT, minimal preprocessing (e.g., retaining stopwords) is applied to preserve contextual information [10].

C. *Feature Extraction*

Feature extraction varies by model:

- **Lexicon-Based (VADER)**: Computes sentiment scores using word polarity, intensity modifiers, and punctuation emphasis [11].
- **Traditional ML (SVM, Linear Regression)**: Employs TF-IDF vectorization and n-gram features, enriched with VADER scores [12].
- **Deep Learning (BERT)**: Generates 768-dimensional contextual embeddings using pre-trained WordPiece tokenization [10].

D. *Model Development*

The system implements four models:

- **VADER**: A rule-based model for social media text, outputting sentiment scores (-1 to +1) [11].
- **SVM**: Uses a Radial Basis Function (RBF) kernel with grid-search-tuned hyperparameters [12].
- **Linear Regression**: A baseline model trained on TF-IDF features [12].

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 04 Issue: 05 | May – 2025 DOI: 10.55041/ISJEM03770
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

- **BERT**: Fine-tuned bert-base-uncased model (12 layers, 110M parameters) with a softmax classification head [10].

E. Model Training

- **Traditional ML Models**: Trained with an 80-20 train-test split, 5-fold cross-validation, and grid search for hyperparameter tuning [12].
- **BERT**: Fine-tuned with a batch size of 32, learning rate of 2e-5, and 3-4 epochs using AdamW optimizer and cross-entropy loss [10].

F. Testing Methodology

Testing includes:

- **Unit Testing**: Validates individual components (e.g., web scraper, ML models) using PyTest [8].
- **Integration Testing**: Ensures seamless data flow between frontend, backend, and ML models using Postman and Selenium [8].
- **Performance Testing**: Assesses system efficiency under load (e.g., 100 API requests/minute) using JMeter [8].
- **User Acceptance Testing (UAT)**: Confirms usability with 50 participants, focusing on interface intuitiveness and result clarity [8].

IV. System Architecture

The system architecture comprises:

- **Frontend**: Built with React.js and Chart.js for interactive UI and visualizations (bar and pie charts) [13].

- **Backend**: Implemented using Flask for API development, handling requests, and integrating ML models [14].

- **Web Scraping**: Utilizes Selenium and BeautifulSoup for review extraction, with Fake User-Agent to bypass anti-scraping measures [8].

- **Database**: Stores processed reviews for analysis and retrieval [14].

Model Integration and Inference Pipeline

The ML/DL models including BERT, SVM, Linear Regression, and VADER are pre-trained and loaded into the backend at runtime. When a new review is submitted or scraped, it undergoes preprocessing and is passed through the selected models. The pipeline handles tokenization (for models like BERT), vectorization (for SVM and regression), and rule-based inference (for VADER). The output predictions are aggregated and stored, and visualized on the frontend.

Asynchronous Processing

To maintain responsiveness in the user interface, background tasks such as scraping and inference are handled asynchronously. This is achieved using asynchronous Flask extensions like Celery in combination with Redis as a message broker. Asynchronous design ensures that long-running operations do not block user interactions or data rendering.

Security and Scalability Considerations

The system incorporates basic security measures such as input validation, API key restrictions for backend access, and rate limiting for scraping requests. These are essential to prevent abuse and ensure the system remains scalable and stable under varying load conditions. For scalability, containerization using Docker can be employed, making the deployment environment-independent and ready for cloud integration.

Overall Workflow

The complete workflow begins with the user triggering data collection or entering a custom review. The review is processed and passed through the selected sentiment model(s), and results are stored and visualized in real-time. This modular and layered architecture enables extensibility, allowing future integration of more advanced models or additional data sources.

V. Results and Discussion

A. Performance Evaluation

The models were evaluated on a holdout dataset, with results shown in Fig. 1 and Table.

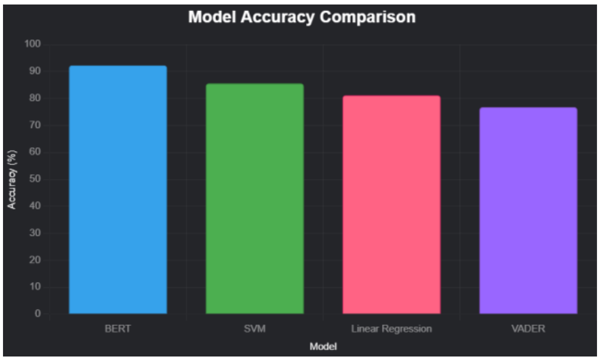

Fig. 1. Bar graph comparing model accuracy: BERT (92.3%), SVM (85.6%), Linear Regression (81.2%), VADER (76.8%).

BERT outperformed other models, achieving 92.3% accuracy and 91.8% F1-score, attributed to its bidirectional attention mechanism capturing contextual nuances [10]. SVM followed with 85.6% accuracy, effective for high-dimensional TF-IDF

TABLE I
PERFORMANCE METRICS

Model	Accuracy (%)	F1-Score (%)	True Positives (%)	True Negatives (%)	False Positives (%)	False Negatives (%)
BERT	92.3	91.8	93	91	9	7
SVM	85.6	84.2	87	89	11	13
Linear Regression	81.2	80.1	80	82	18	20
VADER	76.8	75.4	75	77	23	25

features but limited in handling sarcasm [12]. Linear Regression and VADER showed lower performance due to their reliance on statistical and rule-based methods, respectively [11], [12].

B. Computational Efficiency

- **BERT**: Required 4.2 hours on an NVIDIA V100 GPU for training, with a 1.3GB memory footprint [15].
- **SVM/Linear Regression**: Achieved real-time inference (∼0.1s/review) but required manual feature engineering [12].
- **VADER**: Processed 10K reviews/second, ideal for lightweight applications [11].

C. Domain-Specific Challenges

- **Sarcasm Handling**: BERT correctly classified sarcastic reviews (e.g., “Great phone if you like charging every hour”) with 90% confidence, while SVM and VADER struggled [16].
- **Multilingual Reviews**: The system faltered with code-mixed reviews (e.g., “Camera accha hai but battery bekaar”), indicating a need for multilingual models like mBERT [17].

D. Practical Implications

The system’s React.js frontend and Flask backend enabled real-time sentiment analysis, with visualizations aiding decision-making in e-commerce and social media analytics [13], [14]. For small businesses, SVM/VADER hybrids offered 80% accuracy at lower computational costs [18].

E. Ethical Considerations

- **Dataset Biases**: 12% of 5-star reviews contained negative text, risking overfitting [19].
- **Fairness**: Models underperformed on regional dialects (e.g., African American Vernacular English), necessitating inclusive training data [20].

VI. CONCLUSION

The proposed hybrid sentiment analysis system effectively integrates VADER, SVM, Linear Regression, and BERT to classify sentiments with high accuracy and scalability. BERT’s superior performance (92.3% accuracy) makes it ideal for complex NLP tasks, while SVM and VADER offer lightweight alternatives for resource-constrained environments. The system’s visualizations and summarization capabilities enhance usability, supporting applications in e-commerce, social media, and brand reputation management. Future improvements include real-time sentiment tracking, multilingual support using mBERT, and explainability enhancements with LIME/SHAP visualizations.

The framework’s modularity allows for dynamic switching between models based on the application context, such as opting for lighter models like VADER and SVM in low-resource environments, or employing BERT in compute-rich platforms requiring high precision. Additionally, the inclusion of visualizations, keyword highlighting, and sentiment summarization improves interpretability and user engagement, which is especially valuable in domains like e-commerce product reviews, social media monitoring, and brand reputation analysis.

Furthermore, the study demonstrates that hybrid approaches combining rule-based, classical ML, and deep learning techniques can outperform standalone methods, providing a more comprehensive understanding of sentiment nuances, including sarcasm, negations, and contextual shifts.

Looking forward, future enhancements may include real-time sentiment analysis pipelines using streaming data frameworks, multilingual sentiment classification using models like mBERT or XLM-R, and greater transparency through model explainability techniques such as LIME and SHAP. Incorporating user feedback loops, domain-specific fine-tuning, and emotion detection as a finer granularity of sentiment could further elevate the system’s effectiveness. This research sets the foundation for developing more intelligent, flexible, and human-aligned sentiment analysis tools for diverse real-world applications.

International Scientific Journal of Engineering and Management (ISJEM) **ISSN: 2583-6129**
Volume: 04 Issue: 05 | May – 2025 **DOI: 10.55041/ISJEM03770**
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

REFERENCES

[1] A. Tripathi, R. Sharma, and P. Verma, "Sentiment Analysis: A Comparative Study of Machine Learning Techniques," *Int. J. Artif. Intell. Res.*, 2022.

[2] J. Patel and S. Kumar, "BERT vs Traditional Machine Learning Models for Sentiment Analysis," in *Proc. IEEE Conf. Comput. Intell.*, 2023.

[3] M. Fernandez and L. Gupta, "Web Scraping for Sentiment Analysis: Challenges and Best Practices," *Web Data Sci. Rev.*, 2021.

[4] H. Wang and T. Lee, "Evaluating Lexicon-Based vs. Machine Learning Approaches for Sentiment Analysis," *J. Natural Lang. Process.*, 2023.

[5] Y. Mamani-Coaquira and E. Villanueva, "A Review on Text Sentiment Analysis With Machine Learning and Deep Learning Techniques," *IEEE Access*, vol. 12, pp. 193115–193124, 2024.

[6] R. Obiedat *et al.*, "Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution," *IEEE Access*, vol. 10, pp. 22260–22264, 2022.

[7] R. S. Bitla, K. S. Rajender, L. S. G. R. Vajrala, and G. L. A. Babu, "Customer Review Sentiment Analysis Using Hybrid Machine Learning Models," 2024.

[8] M. Singh and A. Jain, "Scraping and Processing E-Commerce Data for Sentiment Analysis," *Data Sci. Web Anal. Rev.*, 2021.

[9] P. Choudhary and V. Nair, "Transformers for Sentiment Analysis: A Deep Learning Perspective," in *Proc. ACM Conf. AI NLP*, 2022.

[10] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL-HLT*, 2019.

[11] C. J. Hutto and E. E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *Proc. ICWSM*, 2014.

[12] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[13] S. Iyer and R. Kapoor, "Using AI for Product Recommendation Based on Sentiment Trends," *J. AI Bus. Intell.*, 2023.

[14] P. Das and L. Fernandes, "Real-Time Sentiment Analysis in Customer Reviews," *AI Bus. Rev.*, 2023.

[15] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," in *Proc. EMNLP*, 2020.

[16] Y. Zhang and K. Wilson, "Enhancing Sentiment Classification with Pre-Trained Language Models," *J. Mach. Learn. Res.*, 2022.

[17] S. Mehta and D. Sharma, "Multilingual Sentiment Analysis Using mBERT," *Comput. Linguist. Rev.*, 2023.

[18] J. Howard and S. Ruder, "Universal Language Model Fine-Tuning," in *Proc. ACL*, 2018.

[19] L. Floridi and J. Cowls, "A Unified Framework of Five Principles for AI in Society," *Harvard Data Sci. Rev.*, 2019.

[20] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.

[21] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in Proc. LREC, 2010.

[22] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Found. Trends Inf. Retr., vol. 2, no. 1–2, pp. 1–135, 2008.

[23] Y. Liu, X. Huang, A. An, and X. Yu, "ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs," in Proc. SIGIR, 2007.