iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69269

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Sentiment Analysis using BERT, CNN and BI-LSTM

S. Saraswathi[1], Athul K[2], S. Abhinav[3], S. Sivasankar[4]
Department of Information Technology, Puducherry Technological University, Puducherry, India

Abstract: Sentiment Analysis (SA) has emerged as a critical task in Natural Language Processing (NLP), especially with the exponential growth of user-generated content. Although deep learning models like Bi-directional Long Short-TermMemory(Bi-LSTM),Convolutional Neural Networks (CNN), and Bidirectional Encoder Representations from Transformers (BERT) have greatly improved sentiment classification accuracy, they often fail to capture the meaning of idiomatic expressions. In this paper, we propose a hybrid sentiment analysis framework that integrates BERT embeddings with CNN for feature extraction, Bi-LSTM for sequence modeling, and an attention mechanism for enhanced sentiment prediction. Additionally, we introduce a dynamic idiom detection and sentiment inference module that identifies both known and unknown idioms. For unknown idioms, contextual information is retrieved via web search to lexically infer sentiment. The model is evaluated on a benchmark English hotel review dataset, along with an extended idiom-labeled dataset. Our approach achieves a state-of-the-art accuracy of 92.24%, outperforming traditional Bi-LSTM and RNN models. Results show that our hybrid system significantly enhances sentiment classification performance, particularly in the presence of idiomatic expressions.
Keywords: Sentiment analysis, BERT, CNN, Bi-LSTM, idiom detection, attention mechanism, lexicon-based sentiment inference.

I. INTRODUCTION

Sentiment analysis (SA) has emerged as a key task in Natural Language Processing (NLP), playing a vital role in understanding public opinion, user preferences, and feedback across platforms such as product reviews, social media, and customer service interactions. The task involves classifying textual data into categories such as positive, negative, or neutral sentiments. Traditional machine learning approaches, while effective in certain contexts, often struggle with capturing deeper contextual cues, idiomatic expressions, and sequential dependencies within sentences. With the advent of deep learning and transformer-based models, sentiment classification has seen notable improvements in both accuracy and contextual understanding [1].

Role of Deep Learning in Sentiment Analysis

- Transformer Models (e.g., BERT): Pre-trained models like BERT capture deep contextual relationships between words using attention mechanisms, significantly improving classification in complex sentences.
- Sequential Models (Bi-LSTM): Bi-directional Long Short-Term Memory (Bi-LSTM) networks retain long-range dependencies in both forward and backward directions, making them effective in processing sentence structures.
- Convolutional Neural Networks (CNN): CNNs are utilized for feature extraction by detecting local patterns such as n-grams and phrases within text data.

Despite these advances, idiomatic expressions continue to present challenges for sentiment classifiers. Idioms often carry meanings that differ from the literal interpretations of the individual words, leading to misclassifications when using standard models.

A. Challenges in Idiom-Based Sentiment Classification

Idioms such as “kick the bucket” or “spill the beans” convey sentiments not directly inferable from their components. Most sentiment models, including those based on deep learning, tend to overlook these expressions or interpret them literally, thereby reducing overall accuracy. A hybrid model that explicitly detects idioms and infers their sentiment meaning is essential to overcome this limitation. Furthermore, idioms vary significantly in structure and sentiment polarity, requiring dynamic handling rather than static rule-based systems [2].

B. Hybrid Deep Learning Architecture for Sentiment Analysis

This paper proposes a hybrid architecture combining BERT embeddings with CNN and Bi-LSTM layers. BERT captures semantic context, CNN extracts local n-gram patterns, and Bi-LSTM models sequential dependencies. An attention mechanism further enhances the model by focusing on key sentiment-bearing tokens. This architecture is designed to handle nuanced expressions, including idioms, more effectively than standalone models.

C. Idiom Detection and Sentiment Inference Module

To address the idiom interpretation gap, we introduce a two-stage idiom processing pipeline:

- Known Idioms: Detected using cosine similarity with embeddings from a fine-tuned idiom training set.
- Unknown Idioms: Identified and classified through contextual web search. Sentiment is inferred lexically using polarity-bearing keywords from retrieved content.

This dual approach ensures both pre-trained knowledge and real-time inference are leveraged for sentiment interpretation. The idiom module acts as a preprocessing layer that flags idiomatic expressions before they are passed to the sentiment classifier.

D. Dataset and Evaluation

The system is evaluated on an English-language hotel review dataset augmented with idiom-labeled entries. The hybrid model achieves a state-of-the-art accuracy of 92.24%, outperforming traditional Bi-LSTM and RNN baselines. Idiom detection accuracy and sentiment inference quality are also measured independently, demonstrating the effectiveness of our idiom module.

E. Contributions

- A novel hybrid CNN-BiLSTM model with BERT embeddings and attention mechanism for sentiment classification.
- A dynamic idiom detection and inference module that addresses both known and unknown idioms.
- Integration of context-aware sentiment inference through real-time web-based analysis for idioms unseen during training.
- Empirical validation on English-language review data, achieving improved accuracy and interpretability.

II. RELATED WORK

A. Sentiment Classification Using Deep Neural Networks

1) Description: This study investigates the use of deep learning models like CNN and LSTM for sentiment classification tasks.
2) Methodology: CNN is used to extract local syntactic and semantic features, while LSTM captures the sequential dependencies in reviews
3) .Limitations: These models struggle with figurative language and idiomatic expressions that deviate from literal interpretations.
4) Improvement: Incorporating pre-trained embeddings like BERT and attention mechanisms can improve idiom handling and sentiment precision.

B. BERT-Based Sentiment Analysis

1) Description: This work focuses on transformer-based models, particularly BERT, to enhance sentiment classification performance.
2) Methodology: BERT is fine-tuned on sentiment datasets to leverage its deep contextual understanding of language.
3) Limitations: While effective for literal and context-rich text, BERT may still misclassify idioms or domain-specific expressions not present in its training corpus.
4) Improvement: Hybrid models that combine BERT with sequential models and idiom detection mechanisms can overcome these gaps.

C. Idiom-Aware Sentiment Models

1) Description: This research addresses the gap in sentiment analysis models when interpreting idiomatic and figurative expressions.
2) Methodology: Uses lexicon-based methods and idiom dictionaries to detect idioms in sentences and adjust sentiment polarity accordingly.

3) Limitations: Rule-based and lexicon-only methods lack adaptability, perform poorly on unseen idioms, and are language-specific.
4) Improvement: Employing deep learning with dynamic idiom detection and contextual sentiment inference increases robustness and generalization.

D. Attention Mechanisms in Hybrid Sentiment Models

1) Description: Explores the integration of attention mechanisms in neural architectures for better focus on sentiment-bearing terms.
2) Methodology: Attention layers are added to Bi-LSTM and CNN models to prioritize important tokens during classification.
3) Limitations: Attention may still focus on irrelevant tokens if idiomatic expressions are not explicitly identified or modeled.
4) Improvement: Coupling attention with idiom recognition improves interpretability and boosts classification accuracy for idiom-rich content.

E. Idiom Detection Using Semantic Similarity and Contextual Embeddings

1) Description: Investigates idiom recognition through similarity measures in embedding spaces using models like Word2Vec or BERT.
2) Methodology: Idioms are detected by computing cosine similarity between phrases in reviews and idiom datasets.
3) Limitations: Accuracy depends heavily on idiom coverage in the training set and cannot handle unseen idioms effectively.
4) Improvement: Introduce a two-tier idiom detection system with known idiom classification and contextual web-based inference for unknown idioms.

III. BEST TECHNIQUES FORSENTIMENT ANALYSIS

A. BERT(Bidirectional Encoder Representations from Transformers)

1) Description: BERT is a state-of-the-art transformer-based model that understands the contextual meaning of words in both directions (left and right). In sentiment analysis, it excels at identifying sentiment even in complex or ambiguous sentences.
2) Key Features: Captures deep contextual dependencies, handles sarcasm and negation, and provides powerful sentence-level embeddings.
3) Application Area: Widely used in product reviews, customer feedback analysis, and social media sentiment detection.

B. CNN (Convolutional Neural Network)

1) Description: CNNs are effective at extracting local features, such as key phrases or patterns associated with sentiment. In sentiment analysis, CNNs identify n-gram level features like “very good” or “not bad”.
2) Key Features: Efficient in capturing spatial features, robust to word order variations, and computationally fast.
3) Application Area: Ideal for short texts, review snippets, and preprocessing layers in hybrid architectures.

C. Bi-LSTM (Bidirectional Long Short-Term Memory)

1) Description: Bi-LSTM networks process text in both forward and backward directions to capture long-range dependencies. They are highly effective for modeling the sequence of words and detecting sentiment progression across sentences.
2) Key Features: Captures both past and future context, handles variable-length inputs, and is suitable for sequential data.
3) Application Area: Used in sentence-level and document-level sentiment analysis, particularly for long reviews or narratives.

D. Hybrid BERT + CNN + Bi-LSTM Architecture

1) Description: The hybrid model combines the contextual power of BERT with the pattern detection of CNN and the sequential understanding of Bi-LSTM. BERT embeddings are fed into CNN layers to capture local features, followed by Bi-LSTM to learn temporal dependencies.
2) Key Features: Maximizes feature extraction and classification accuracy by leveraging strengths of all three models.
3) Application Area: Applied in domains requiring high precision sentiment analysis, such as tourism reviews, financial opinions, and public sentiment tracking.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

IV. TECHNIQUES USED FOR SENTIMENT ANALYSIS

Sentiment analysis involves extracting subjective information from text to determine the emotional tone. In this study, we implement a hybrid deep learning approach that leverages BERT for embeddings, CNN for pattern extraction, and Bi-LSTM for sequential modeling. This fusion helps in capturing both contextual meaning and structural flow of sentiment-bearing expressions.

A. BERT for Contextual Embeddings

BERT generates high-quality word embeddings by analyzing the full sentence context. In our model, BERT is used to encode input text into dense vector representations that are sensitive to word usage and semantics.

Advantages:

- Captures deep semantic relationships
- Pre-trained on large corpora, reducing data needs
- Handles complex grammar and idioms well

B. CNN for Local Pattern Extraction

The CNN layer is applied to BERT embeddings to identify crucial local patterns like sentiment phrases ("very happy", "extremely bad") that may influence overall polarity.

Advantages:

- Fast and efficient feature detection
- Good for short-term dependencies and sentiment expressions
- Complements BERT with local focus

C. Bi-LSTM for Sequence Modeling

The Bi-LSTM layer processes the CNN output in both forward and backward directions, understanding the sentiment flow across the review.

Advantages:

- Captures long-term sentiment dependencies
- Works well on long reviews and structured content
- Robust to context shifts and sentiment changes

D. Sentiment Classification Layer

The features from Bi-LSTM are passed through fully connected layers for final sentiment classification (Negative, Neutral, Positive). Softmax activation ensures probabilistic output for multi-class classification.

Advantages:

- End-to-end deep learning
- Accurate sentiment labeling
- Scalable to multiple domains

V. PROPOSED WORK

The proposed work focuses on enhancing sentiment analysis by incorporating idiom detection and sentiment inference into a hybrid deep learning architecture. Traditional models often misinterpret idiomatic expressions, which can carry non-literal sentiments. To overcome this, the system integrates a hybrid sentiment model with an idiom-aware processing pipeline that identifies and interprets idioms—especially previously unseen ones.

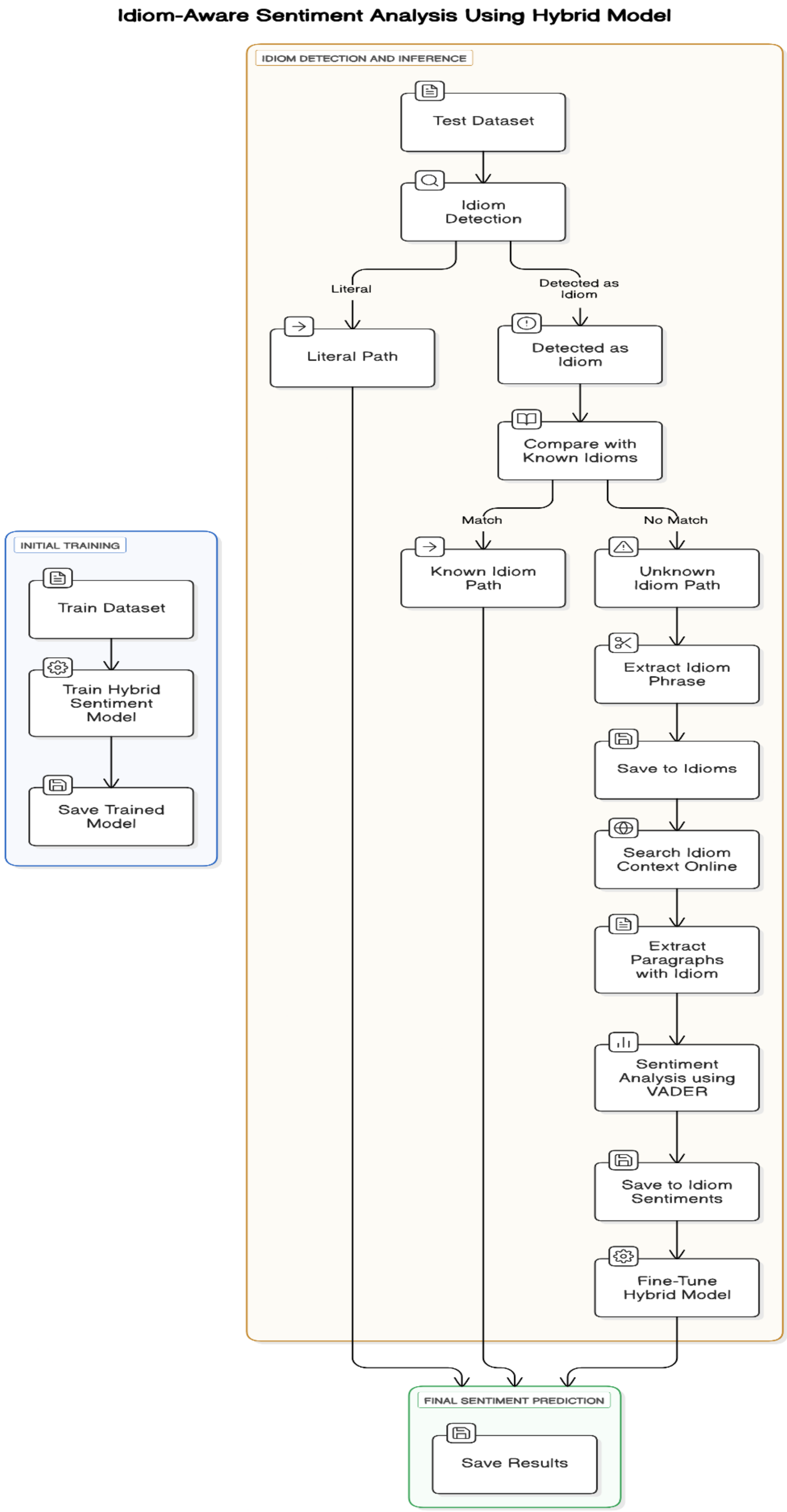

Fig 5.1 Proposed System Design

A. Data Collection

The primary dataset comprises 40,000 English-language hotel reviews sourced from publicly available online platforms. These reviews capture a range of customer experiences related to hotel amenities, staff interaction, cleanliness, and overall satisfaction. Each review is manually annotated with one of three sentiment labels—Positive, Negative, or Neutral—providing a reliable foundation for supervised sentiment analysis.

To enable idiom-aware sentiment modeling, an additional idiom dataset consisting of 200 English idioms annotated with sentiment labels is integrated. These idioms represent common figurative expressions observed in natural language hotel reviews. Furthermore, a test set of 25 English reviews is compiled to evaluate the system's capability to detect idioms and accurately infer sentiment, especially for idioms not encountered during training.

B. Data Preprocessing

A rigorous preprocessing pipeline is applied to ensure data cleanliness, model readiness, and semantic accuracy. The following steps are executed:

1) Text Cleaning: Removal of HTML tags, special characters, punctuation, and emojis to standardize raw text.
2) Tokenization & Lowercasing: Reviews are tokenized and converted to lowercase to normalize input.
3) Stopword Removal & Lemmatization: Irrelevant stopwords are removed and tokens are lemmatized to their root forms.
4) BERT Tokenization: All reviews are tokenized using BERT's WordPiece tokenizer to prepare them for embedding generation.
5) Vector Representation: For idiom comparison, BERT embeddings are used to compute cosine similarity between idiom phrases.

For exploratory analysis, Matplotlib, Seaborn, and Plotly are used to visualize sentiment distributions, idiom frequencies, and correlation heatmaps. This pipeline ensures the dataset is clean, context-aware, and optimized for both idiom detection and sentiment analysis tasks.

C. Idiom Detection and Sentiment Classification Models

The system consists of three key components, each responsible for handling idiomatic and literal reviews efficiently:

1) Idiom Detection Module

A fine-tuned BERT + RoBERTa ensemble model classifies reviews as either Idiom or Literal.
If an idiom is detected:

- The idiomatic phrase is extracted using context-aware matching techniques.
- BERT embeddings are used to compare with known idioms using cosine similarity.
- If similarity > 0.9 → classified as Known Idiom.
- Otherwise → classified as Unknown Idiom and routed to sentiment inference.

2) Sentiment Inference for Unknown Idioms

For idioms not found in the training data:

- The idiom is queried on the web, and relevant text is extracted using newspaper3k.
- Sentiment is inferred from the extracted content using VADER, a lexicon-based sentiment analyzer.
- The inferred idiom and sentiment are stored for future fine-tuning.

3) Hybrid Sentiment Classification Model

A hybrid neural architecture is implemented:

- BERT for contextual word embeddings.
- CNN layers to capture local n-gram patterns.
- Bi-LSTM layers for understanding sequential dependencies.
- Fully Connected Output Layer for sentiment classification.

Initially trained on literal reviews, the model is later fine-tuned using known and inferred idiomatic expressions, enhancing its idiom-aware sentiment understanding.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Table 5.1: Key Features in the Sentiment and Idiom Dataset

Feature Name	Description
Review_Text	Full English review written by the user
Sentiment_Label	Manually annotated sentiment class (Positive, Negative, Neutral).
Contains_Idiom	Binary indicator (1 or 0) showing if the review contains an idiom
Idiom_Expression	Extracted idiomatic phrase from the review (if present).
Idiom_Type	Indicates if the idiom is "Known" (in training set) or "Unknown."
Inferred_Sentiment	Sentiment assigned to unknown idioms using lexicon-based inference.

This multi-level dataset supports robust analysis by encompassing both literal and figurative language expressions in sentiment classification.

D. Model Testing and Evaluation

Model performance is evaluated across both idiom detection and sentiment classification components using a dedicated test set.

1) Evaluation Metrics:

- Accuracy: Overall correctness of predictions.
- Precision: Correctness of positive predictions.
- Recall: Ability to capture all relevant positive instances.
- F1-Score: Harmonic mean of Precision and Recall.
- Confusion Matrix: Detailed view of correct and incorrect classifications.
- ROC-AUC: For idiom classification, measures the model's ability to distinguish between classes.

2) Separate evaluations are conducted for:

- Idiom Detection Accuracy (Idiom vs. Literal)
- Idiom Type Classification Accuracy (Known vs. Unknown)
- Sentiment Classification Accuracy with and without idioms

This multi-level evaluation framework ensures that the system performs robustly in identifying idioms, distinguishing between known and unknown cases, and classifying sentiment with high precision in both literal and idiomatic review contexts.

Table 5.2 Libraries used in Implementing and Evaluating the Performance of the Model

Library	Purpose
numpy	Supports large, multi-dimensional arrays and matrices, and provides a collection of mathematical functions.
pandas	Facilitates data manipulation and analysis, offering data structures like DataFrames.
joblib	Efficiently saves and loads large Python objects, such as trained machine learning models.

sklearn.metrics	Offers functions for evaluating models, including accuracy, precision, recall, F1-score, and mean squared error (MSE).
sklearn.model_selection	Provides tools for splitting datasets, cross-validation, and hyperparameter tuning.
newspaper3k	Extracts textual content from URLs for real-world idiom usage contexts.
duckduckgo_search	Performs web search to collect real-world idiom examples from online articles.
stanza	Used for stopword removal, lemmatization, and POS tagging (especially in English text preprocessing).

VI. RESULTS

This section presents the performance of the sentiment analysis model before and after fine-tuning it with inferred sentiment labels for previously unknown idioms. Evaluation metrics include accuracy, precision, recall, F1-score, and the confusion matrix.

A. Initial Performance Before Fine-Tuning

The model was initially evaluated on idiomatic test data without being trained on unknown idioms. The results demonstrated poor generalization, particularly for figurative language.

1) Sample Misclassifications

Review: Room decor was easy on the eyes.

- Labeled Sentiment: Positive
- Predicted Sentiment: Negative
- Review: They patched things up after the mishap.
- Labeled Sentiment: Positive
- Predicted Sentiment: Negative

These samples illustrate the model's struggle with understanding idiomatic expressions.

Label	Precision	Recall	F1-Score	Support
Negative	0.50	1.00	0.67	34
Positive	1.00	0.06	0.11	36
Accuracy			0.5143	70
Macro Avg	0.75	0.53	0.39	70
Weighted Avg	0.76	0.51	0.38	70

B. Improved Performance After Fine-Tuning with Inferred Idioms

To enhance idiom understanding, inferred sentiment labels were generated for previously unknown idioms through web-based context retrieval and VADER sentiment analysis. These newly labeled idioms were used to fine-tune the model.

Classification Report (After Fine-Tuning)

Label	Precision	Recall	F1-Score	Support
Negative	1.00	0.89	0.94	9
Positive	0.93	1.00	0.96	13
Accuracy			0.9545	22
Macro Avg	0.96	0.94	0.95	22
Weighted Avg	0.96	0.95	0.95	22

C. Performance Comparison: Before vs. After Fine-Tuning

Metric	Before Fine-Tuning	After Fine-Tuning
Accuracy	51.43%	95.45%
Positive Precision	1.00	0.93
Positive Recall	0.06	1.00
Positive F1-Score	0.11	0.96
Macro Avg F1-Score	0.39	0.95
Weighted Avg F1-Score	0.38	0.95

VII. CONCLUSION

Based on the comparative analysis conducted before and after fine-tuning, it can be concluded that the integration of idiom detection and sentiment inference significantly enhanced the overall performance of the sentiment analysis model.

Prior to fine-tuning, the hybrid model (BERT + CNN + Bi-LSTM) exhibited poor generalization on idiomatic expressions, with an overall accuracy of only 51.43% and a positive class recall of just 0.06, indicating its inability to correctly identify sentiment when idioms were present. This weakness highlighted the model’s lack of exposure to figurative language during training.

After incorporating unknown idioms through a detection and inference pipeline—where sentiments were inferred via contextual web scraping and lexicon-based analysis—the model was fine-tuned on the extended dataset. Post-fine-tuning evaluation showed a remarkable improvement, with the model achieving an accuracy of 95.45%, a perfect recall for positive sentiments (1.00), and an overall F1-score of 0.95.

These results demonstrate that:

1) Idiom detection is essential for real-world sentiment analysis tasks, particularly in domains like hotel reviews where figurative language is prevalent.
2) Sentiment inference for unknown idioms can be reliably performed using contextual search and lexicon-based methods.
3) Fine-tuning with the newly inferred idioms transforms the model from being overly literal to effectively understanding figurative speech.

Thus, the final hybrid model proves to be robust, context-aware, and capable of accurately handling both literal and idiomatic reviews, establishing it as a reliable approach for sentiment analysis in real-world applications.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

REFERENCES

[1] L. Khan, A. Qazi, H.-T. Chang, M. Alhajlah, and A. Mahmood, "Empowering Urdu sentiment analysis: an attention-based stacked CNN-Bi-LSTM DNN with multilingual BERT," Complex & Intelligent Systems, vol. 11, pp. 10, 2025. [Online]. Available: https://doi.org/10.1007/s40747-024-01631-9.

[2] Wouter van Atteveldt, Mariken A. C. G. van der Velden & Mark Boukes,"The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms",Communication Methods and Measures, 28 Jan 2021.

[3] Nhan Cach Dang, María N. Moreno-García, Fernando De la Prieta, "Sentiment Analysis Based on Deep Learning: A Comparative Study", Electronics, Volume 9, Issue 4, April 2020, ISSN 2079-9292.

[4] Santwana Sagnika, Anshuman Pattanaik, Bhabani Shankar Prasad Mishra, Saroj K. Meher, "A Review on Multi-Lingual Sentiment Analysis by Machine Learning Methods", Journal Articles, Indian Statistical Institute, 2020.

[5] George Manias, Argyro Mavrogiorgou, Athanasios Kiourtis, Chrysostomos Symvoulidis, Dimosthenis Kyriazis, "Multilingual Text Categorization and Sentiment Analysis: A Comparative Analysis of the Utilization of Multilingual Approaches for Classifying Twitter Data", Neural Computing and Applications, Volume 35, Issue 21415–21431, May 2023, DOI: 10.1007/s00521-023-08629-3.

[6] Adwan OY, Al-Tawil M, Huneiti AM, Shahin RA, Abu Zayed AA, Al-Dibsi RH (2020) Twitter sentiment analysis approaches: a survey. Int J Emerg Technol Learn. https://doi.org/10.3991/ijet.v15i15.14467