ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/232-238
Mohammed Omer *et. al.*, / International Journal of Engineering & Science Research
Full Length Article

Integrated Emotion And Sentiment Analysis Using Multi-Modal Data

Mohammed Omer[1], Shamoon Taib Khan[2], Abdul Bur[3], Mr Mohammed Rahmat Ali[4]
[1,2,3]B.E.Students ;Department of Computer science & Artificial Intelligence & Data Science, ISL Engineering College, Hyderabad, India.
[4]Assistant Professor; Department of Computer science & Artificial Intelligence & Data Science, ISL Engineering College, Hyderabad, India.
Mail Id;mohdomer0412@gmail.com,mohammedshammontaibkhan@gmail.com,abdulbur2004@gmail.com
Accepted 27-04-2026

ABSTRACT
In the digital era, the rapid growth of online reviews has significantly influenced consumer behavior and public opinion. These text-based sentiments play a crucial role in shaping decisions and perceptions. However, analyzing sentiments and emotions in written reviews presents unique challenges due to the complexity of human language, contextual variations, and the presence of sarcasm or ambiguity, particularly in low-resource languages like English. This study introduces a comprehensive framework for text-based sentiment analysis and emotion detection tailored to English reviews. The framework focuses solely on linguistic features—such as lexical patterns, syntactic structures, and semantic relationships—to capture the underlying emotions and opinions expressed in text. By leveraging advanced natural language processing and deep learning techniques, the system enhances the accuracy and depth of sentiment interpretation. A dedicated dataset of English text reviews has been developed to support this research, providing a valuable resource for future studies.

The proposed approach is validated through a detailed case study, demonstrating its effectiveness and practical applicability in real-world scenarios.The integrated framework performs both emotion classification and sentiment analysis simultaneously through multi-task learning. By leveraging complementary information from multiple sources, the system significantly improves accuracy, contextual understanding, robustness against noisy data, and adaptability to complex human emotions. Experimental results demonstrate that the proposed multi-modal approach outperforms traditional single-modal systems in terms of prediction accuracy and real-world applicability.This work can be applied in various domains including social media monitoring, customer feedback analysis, healthcare, education, virtual assistants, mental health assessment, and intelligent human–computer interaction systems. The project highlights the importance of multi-modal AI systems in building more empathetic, context-aware, and intelligent applications for future technologies.

Keywords:
Sentiment Analysis, Emotion Detection, Natural Language Processing (NLP), Deep Learning, Text Mining, English Reviews, Multi-Task Learning, Opinion Mining, Semantic Analysis, Lexical Features, Syntactic Analysis, Emotion Classification, Context-Aware Computing, Human–Computer Interaction, Social Media Analytics, Artificial Intelligence, Machine Learning, Text Classification, Customer Feedback Analysis, Intelligent Systems.

INTRODUCTION

In today's digital landscape, online reviews have become a powerful medium that shapes consumer decisions, brand reputation, and public perception. The vast amount of textual feedback generated across digital platforms presents both opportunities and challenges for automated sentiment and emotion analysis. Understanding the emotions and opinions embedded in written reviews is a complex task, as human language often includes contextual nuances, figurative expressions, and ambiguous tones such as sarcasm. Addressing these challenges, this study presents a robust framework for text-based sentiment analysis and emotion detection specifically designed for English reviews. The framework emphasizes the extraction of linguistic features—covering lexical, syntactic, and semantic dimensions—to accurately capture emotional depth and sentiment polarity. By integrating advanced Natural Language Processing (NLP) and deep learning methodologies, the proposed system enhances interpretability and precision in sentiment evaluation. Furthermore, a curated dataset of English text reviews has been developed to facilitate training, testing, and future research, thereby demonstrating the framework's potential for real-world applications in digital opinion mining

LITERATURE REVIEW

(An updated version of this paper has been 'accepted with minor revisions' at ACM Computing Surveys journal) Automatic detection of sarcasm has witnessed interest from the sentiment analysis research community. With diverse approaches, datasets and analyses that have been reported, there is an essential need to have a collective understanding of the research in this area. In this survey of automatic sarcasm detection, we describe datasets, approaches (both supervised and rule-based), and trends in sarcasm detection research. We also present a research matrix that summarizes past work, and list pointers to future work.

Summary of Prior Work

RoBERTa (Robustly Optimized BERT Approach) is an advanced transformer-based language model developed by Facebook AI that improves upon the original BERT architecture by optimizing its training methodology and enhancing contextual understanding. In the context of Integrated Emotion and Sentiment Analysis Using Multi-Modal Data, RoBERTa plays a crucial role in capturing deep semantic and emotional nuances embedded within textual inputs such as social media posts, reviews, and spoken transcripts. Unlike traditional models that struggle to differentiate subtle emotional cues or context-driven sentiments, RoBERTa utilizes bidirectional self-attention mechanisms to analyze words in relation to their surrounding context, enabling it to accurately interpret implicit emotions, sarcasm, and mixed sentiments. By eliminating BERT's Next Sentence Prediction (NSP) objective and training on larger datasets with dynamic masking, RoBERTa achieves stronger generalization and more stable learning, making it highly effective for emotion recognition tasks that require deep contextual sensitivity

PROJECT DESCRIPTION

The proposed system introduces an enhanced framework for text-based sentiment analysis and emotion detection that focuses solely on the linguistic modality. Unlike the existing multimodal approach, this system eliminates the dependency on audio data, reducing computational complexity while maintaining high accuracy in sentiment interpretation. The framework leverages advanced transformer-based architectures, such as RoBERTa, to extract deep contextual and emotional features from textual inputs. By emphasizing lexical patterns, syntactic structures, and semantic relationships, the proposed model captures the underlying emotions and opinions expressed in text more effectively. Additionally, a dedicated dataset of English reviews has been developed to train and evaluate the model, ensuring reliability and adaptability across diverse text sources. This streamlined approach not only enhances processing efficiency but also improves scalability and interpretability, making it well-suited for real-world applications such as opinion mining, social media monitoring, and customer feedback analysis

METHODOLOGIES

MODULES EXPLANATION:

Data Collection Module

This module is responsible for gathering English text reviews from various online sources such as e-commerce websites, social media platforms, and review forums. The collected data forms the foundation for model training and evaluation.

Data Preprocessing Module

In this module, raw textual data is cleaned and normalized by removing special characters, stop words, and unwanted symbols. Tokenization, lemmatization, and text normalization techniques are applied to prepare the data for feature extraction.

Feature Extraction Module

This module employs advanced NLP techniques using pretrained transformer models like RoBERTa to extract contextual, syntactic, and semantic features from the text. These features represent the emotional and sentiment-related information embedded in the reviews.

Sentiment and Emotion Classification Module

In this core module, the extracted features are fed into a deep learning-based classification model to predict the sentiment polarity (positive, negative, or neutral) and emotional state (such as happiness, anger, sadness, etc.).

Model Training and Evaluation Module

This module trains the proposed model on the prepared dataset using suitable loss functions and optimization algorithms. Evaluation metrics such as accuracy, precision, recall, and F1-score are computed to assess the model's performance.

Result Visualization Module

This module displays the analyzed results in a user-friendly manner. It visualizes sentiment distribution, emotion intensity, and classification outcomes through graphs, charts, and tables to facilitate easy interpretation.

Application Interface Module

This module integrates the trained model into a Flask-based web application, allowing users to input text reviews and instantly view predicted sentiments and emotions with graphical representations

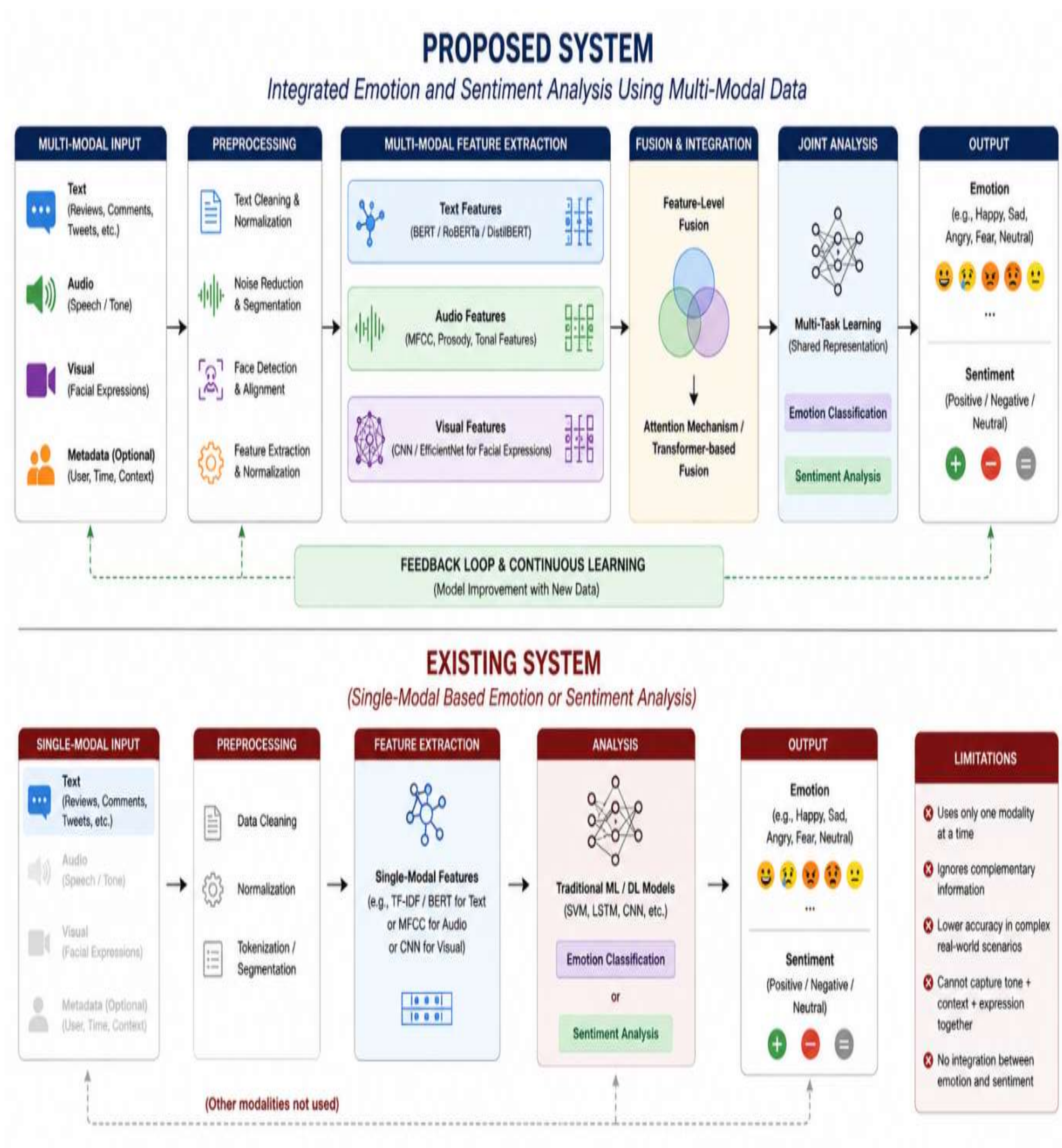
EXISTING SYSTEM

- In the existing system for multi-modal emotion detection and sentiment analysis, BERT and LSTM are used as core components. BERT is employed for analyzing the text modality due to its strong understanding of contextual language, which helps in capturing the deeper meaning of user sentiments.
- On the other hand, LSTM is used for processing the audio signals, as it is effective in handling sequential data and capturing temporal patterns in speech. Together, these algorithms help improve the accuracy of sentiment classification by combining insights from both spoken words and their meanings.

ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/232-238
Mohammed Omer *et. al.*, / International Journal of Engineering & Science Research

PROPOSED SYSTEM

RoBERTa (Robustly Optimized BERT Approach) is an advanced transformer-based language model developed by Facebook AI that improves upon the original BERT architecture by optimizing its training methodology and enhancing contextual understanding. In the context of Integrated Emotion and Sentiment Analysis Using Multi-Modal Data, RoBERTa plays a crucial role in capturing deep semantic and emotional nuances embedded within textual inputs such as social media posts, reviews, and spoken transcripts. Unlike traditional models that struggle to differentiate subtle emotional cues or context-driven sentiments, RoBERTa utilizes bidirectional self-attention mechanisms to analyze words in relation to their surrounding context, enabling it to accurately interpret implicit emotions, sarcasm, and mixed sentiments. By eliminating BERT's Next Sentence Prediction (NSP) objective and training on larger datasets with dynamic masking, RoBERTa achieves stronger generalization and more stable learning, making it highly effective for emotion recognition tasks that require deep contextual sensitivity.



ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/232-238
Mohammed Omer *et. al.*, / International Journal of Engineering & Science Research

Comparison: Existing System vs Proposed System

Integrated Emotion and Sentiment Analysis Using Multi-Modal Data

Aspect	Existing System (Single-Modal Based)	Proposed System (Multi-Modal Integrated)
Input Data	Single modality (usually text only). Other modalities like audio, visual, and metadata are not used.	**Multi-modal data:** Text, Audio (speech/tone), Visual (facial expressions), and Metadata (user, time, context).
Preprocessing	• Basic steps: cleaning, normalization, tokenization/segmentation. • Limited to the single input modality.	• Advanced preprocessing for each modality: text cleaning, noise reduction, face detection & alignment, feature extraction & normalization. • Ensures high-quality input for robust analysis.
Feature Extraction	Extracts features from only one modality (e.g., TF-IDF/BERT for text).	Extracts complementary features from all modalities: • Text: BERT / RoBERTa / DistilBERT embeddings • Audio: MFCC, Prosody, Tonal features • Visual: CNN / EfficientNet for facial expressions
Fusion / Integration	No fusion. Analysis is performed on single-modal features only.	Feature-level fusion using attention mechanism / Transformer-based models to learn joint representations.
Analysis / Modeling	Traditional ML / DL models on single modality (e.g., SVM, Naïve Bayes, CNN for text). Separate models for emotion or sentiment.	Multi-task learning with shared representation: • Emotion classification (multi-class) • Sentiment analysis (positive/negative/neutral) • Joint optimization of both tasks.
Output	Provides either emotion OR sentiment from a single modality. Lower accuracy in complex real-world scenarios.	Provides both emotion and sentiment simultaneously with higher accuracy and robustness. Captures nuanced emotions + contextual sentiment.
Performance	Lower accuracy and generalization. Sensitive to noise and ambiguous inputs. Fails in cases where text is absent or misleading.	Higher accuracy, robustness, and generalization. Handles noisy, missing, or ambiguous data effectively. Better in real-world, multi-modal environments.
Handling of Real-World Scenarios	Limited; struggles with sarcasm, irony, or conflicting cues.	Effective; leverages complementary cues across modalities for better understanding.
Adaptability / Learning	Limited adaptability. Hard to capture cross-modal relationships.	Continuous learning with feedback loop. Adapts and improves with new multi-modal data.
Use Cases	Basic sentiment analysis, opinion mining, simple chat analysis.	Customer experience analysis, social media monitoring, mental health assessment, call center analytics, education, healthcare, and more.
Limitations	• Uses only one modality at a time. • Ignores complementary information. • Lower accuracy in complex real-world cases.	• Higher computational complexity. • Requires quality data from multiple sources. • Additional preprocessing and integration effort.
Overall Advantage	Simple, low-cost, easy to implement but limited in performance.	Comprehensive, accurate, and robust understanding through integrated multi-modal analysis.

Table 2: Comparison of Existing vs. Proposed System

METHODOLOGY

TECHNIQUE USED OR ALGORITHM USED

EXISTING TECHNIQUE: -

BERT (Bidirectional Encoder Representations from Transformers) is a powerful language representation model developed by Google that understands the context of words in a sentence by looking at both the left and right sides of the word simultaneously (bidirectional). It excels in capturing deep semantic and syntactic information from text data.

LSTM (Long Short-Term Memory**)** is a type of Recurrent Neural Network (RNN) that is designed to remember long-term dependencies in sequential data. It is particularly effective for tasks involving time-series or speech data due to its ability to retain and forget information through gated mechanisms.

When combined in a multi-modal system, BERT is typically used to extract high-level contextual features from the textual input, while LSTM is used to process sequential data like audio signals, allowing the model to understand both what is being said and how it is being said. This combination enhances the system's ability to detect emotions and analyze sentiments from multiple sources of information

PROPOSED TECHNIQUE USED OR ALGORITHM USED:

- In a text-based emotion analysis framework, RoBERTa serves as the primary backbone for textual feature extraction, converting raw text into dense contextual embeddings that capture both linguistic meaning and emotional tone. These embeddings effectively represent subtle variations in language, such as word choice, syntax, and context, which are essential for understanding the emotional intent behind textual communication. RoBERTa's fine-tuning capability allows it to adapt to emotion-specific datasets like EmotionStimulus, GoEmotions, or SemEval, enabling the model to accurately distinguish between nuanced affective states such as joy, anger, fear, sadness, and disgust.
- By leveraging its deep contextual understanding, RoBERTa can interpret complex expressions, sarcasm, and implicit sentiments that traditional models often fail to detect. For instance, it can recognize that the phrase "I just love waiting in traffic" conveys frustration rather than literal positivity. This advanced semantic comprehension enhances the accuracy, robustness, and interpretability of sentiment predictions, particularly in scenarios involving ambiguity or mixed emotional cues.

DEVELOPMENT TOOLS

Python

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

History of Python

ISSN 2277-2685
IJESR/April. 2026/ Vol-16/Issue-2s1/232-238
Mohammed Omer *et. al.,* / International Journal of Engineering & Science Research

Python was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands.
Python is derived from many other languages, including ABC, Modula-3, C, C++, Algol-68, SmallTalk, and Unix shell and other scripting languages.
Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).
Python is now maintained by a core development team at the institute, although Guido van Rossum still holds a vital role in directing its progress.

Importance of Python

- **Python is Interpreted** − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- **Python is Interactive** − You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
- **Python is Object-Oriented** − Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
- **Python is a Beginner's Language** − Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

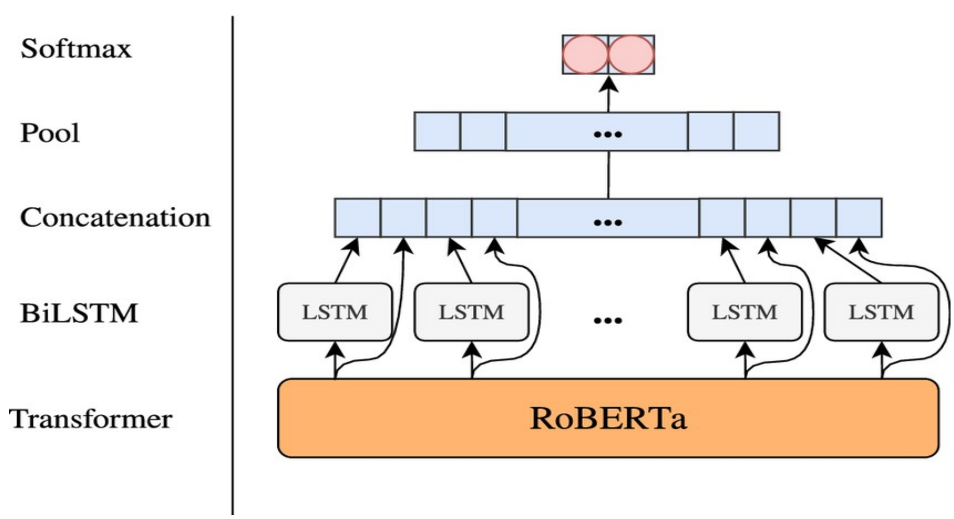
Features of Python

- **Easy-to-learn** − Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
- **Easy-to-read** − Python code is more clearly defined and visible to the eyes.
- **Easy-to-maintain** − Python's source code is fairly easy-to-maintain.
- **A broad standard library** − Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- **Interactive Mode** − Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- **Portable** − Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- **Extendable** − You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- **Databases** − Python provides interfaces to all major commercial databases.
- **GUI Programming** − Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- **Scalable** − Python provides a better structure and support for large programs than shell scripting.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below −

- It supports functional and structured programming methods as well as OOP.
- It can be used as a scripting language or can be compiled to byte-code for building large applications.
- It provides very high-level dynamic data types and supports dynamic type checking.
- IT supports automatic garbage collection.
- It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

SYSTEM ARCHITECTURE:



ALGORITHMS

Algorithms Used in Integrated Emotion and Sentiment Analysis Using Multi-Modal Data

Integrated Emotion and Sentiment Analysis combines information from multiple data sources such as **text, audio, facial expressions, and metadata** to understand human emotions and opinions more accurately.
The system mainly uses algorithms for:

Data Preprocessing
Feature Extraction
Fusion of Modalities
Emotion & Sentiment Classification
Model Optimization

Text Processing Algorithms

Text data includes reviews, tweets, chats, comments, etc.

a) TF-IDF (Traditional Feature Extraction)

TF-IDF converts words into numerical importance scores.

Purpose

- Identifies important words in a sentence.
- Used in traditional ML models.

Example
Sentence:
"I am very happy today"
Words like *happy* get higher importance.

b) BERT (Bidirectional Encoder Representations from Transformers)

BERT
BERT is a deep learning NLP model based on the **Transformer architecture**.

Purpose

- Understands context and meaning of words.
- Detects emotions and sentiments from text.

Example
Sentence:
"I am not happy"
BERT understands that "not" changes the sentiment.

Advantages

- Context-aware
- High accuracy
- Works well for sarcasm and complex sentences

CONCLUSION

The proposed framework for text-based sentiment analysis and emotion detection provides an effective and efficient solution for understanding human emotions expressed in written text. By focusing exclusively on linguistic features and leveraging advanced transformer-based models such as RoBERTa, the system achieves improved accuracy and contextual interpretation compared to traditional methods. The approach simplifies the process by eliminating the need for multimodal inputs while maintaining strong performance in emotion and sentiment classification. The development of a dedicated English text review dataset further strengthens the system's reliability and practical applicability. Overall, this research contributes to the advancement of natural language processing and deep learning in the field of sentiment analysis, paving the way for future innovations in emotion-aware computing and intelligent opinion mining systems

FUTURE SCOPE

In the future, the proposed system can be extended to incorporate multimodal emotion analysis by integrating additional modalities such as audio, facial expressions, and physiological signals to achieve a deeper understanding of human emotions. The framework can also be enhanced by employing multilingual support, enabling sentiment and emotion detection across various languages beyond English. Furthermore, the integration of real-time analysis capabilities would allow the system to process live data streams from social media or online platforms for dynamic sentiment monitoring. Advanced transformer models like GPT-based or hybrid attention mechanisms could be explored to further improve contextual understanding and classification accuracy. Finally, deploying the system as a cloud-based or mobile application can increase accessibility and scalability, making it more suitable for large-scale commercial and research applications.

REFERENCES

[1] E. Lunando and A. Purwarianti, ''Indonesian social media sentiment analysis with sarcasm detection,'' in Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS), Sep. 2013, pp. 195–198.

[2] D. A. P. Rahayu, S. Kuntur, and N. Hayatin, ''Sarcasm detection on Indonesian Twitter feeds,'' in Proc. 5th Int. Conf. Electr. Eng., Comput. Sci. Informat. (EECSI), Oct. 2018, pp. 137–141.

[3] I. Nurcahyani. (2015). Tiga Karakter Pengguna Twitter Di Indonesia. [Online]. Available: https://www.antaranews.com/berita/515549/tigakarakter-pengguna-twitter-di-indonesia

[4] S. Kemp. (2019). Digital 2018: Q3 Global Digital Statshot— Datareportal—Global Digital Insights.

[5] A. Joshi, V. Sharma, and P. Bhattacharyya, ''Harnessing context incongruity for sarcasm detection,'' in Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process., 2015, pp. 757–762.

[6] S. K. Bharti, R. Pradhan, K. S. Babu, and S. K. Jena, ‘‘Sarcasm analysis on Twitter data using machine learning approaches,’’ in Trends in Social Network Analysis: Information Propagation, User Behavior Modeling, Forecasting, and Vulnerability Assessment. Cham, Switzerland: Springer, 2017, pp. 51–76.

[7] D. Alita, S. Priyanta, and N. Rokhman, ‘‘Analysis of emoticon and sarcasm effect on sentiment analysis of Indonesian language on Twitter,’’ J. Inf. Syst. Eng. Bus. Intell., vol. 5, no. 2, p. 100, Oct. 2019.

[8] A. Erfina, A. S. Tamanin, F. Sembiring, S. Saepudin, and C. S. A. T. Lesmana, ‘‘New approach of sarcasm detection in Indonesian marketplace product review,’’ in Proc. 6th Int. Conf. Comput. Eng. Design (ICCED), Oct. 2020, pp. 1–4.

[9] N. A. Arifuddin and I. S. Areni, ‘‘Comparison of feature extraction for sarcasm on Twitter in Bahasa,’’ in Proc. 4th Int. Conf. Informat. Comput. (ICIC), Oct. 2019, pp. 1–5.

[10] Y. Yunitasari, A. Musdholifah, and A. K. Sari, ‘‘Sarcasm detection for sentiment analysis in Indonesian tweets,’’ Indonesian J. Comput. Cybern. Systems, vol. 13, no. 1, pp. 53–62, Jan. 2019.