ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

INTELLIGENT EMOTION RECOGNITION AND SENTIMENT CLASSIFICATION FROM TEXT USING ROBERTa

M. RAMAN KUMAR [1], DONTHULA THRISHIKA [2], GUNDEBOYINA ANUSHA [3], GUMPENA YASHWANTH [4], ENIGA NITHINDHAR REDDY [5]

[1]Assistant Professor, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

[2, 3, 4, 5] B.Tech, Dept. of CSE, Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India

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.

Index Terms – Emotion Recognition, Sentiment Classification, RoBERTa, Natural Language Processing, Transformer Models,

Deep Learning, Sentiment Analysis, Text Classification, Machine Learning, Artificial Intelligence.

I. 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.

II. RELATED WORK

A. A survey of automatic sarcasm detection: Fundamental theories, formulation, datasets, detection methods, and opportunities

(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.

B. Arabic sarcasm detection: An enhanced fine-tuned language model approach

This research paper addresses the significant challenge of sarcasm detection and sentiment analysis within the context of Arabic tweets. It highlights the efficacy of the "AraBert-Arabic-Sentiment-Analysis" model, a pre-trained model that demonstrates robust performance in these tasks. When fine-tuned, this model achieves an accuracy of 87.8% in sarcasm detection and 77.4% in sentiment analysis. Our findings also suggest that extensive preprocessing of data is not invariably beneficial. The optimal results were observed with minimal preprocessing, underscoring the importance of retaining critical features in the data. This study emphasizes the need for a balanced approach to data preprocessing to enhance the effectiveness of computational models in handling nuanced linguistic tasks.

C. One country, 700+ languages: NLP challenges for underrepresented languages and dialects in Indonesia

NLP research is impeded by a lack of resources and awareness of the challenges presented by underrepresented languages and dialects. Focusing on the languages spoken in Indonesia, the second most linguistically diverse and the fourth most populous nation of the world, we provide an overview of the current state of NLP research for Indonesia's 700+ languages. We highlight challenges in Indonesian NLP and how these affect the performance of current NLP systems. Finally, we provide general recommendations to help develop NLP technology not only for languages of Indonesia but also other underrepresented languages.

III. PROPOSED METHODOLOGY

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.

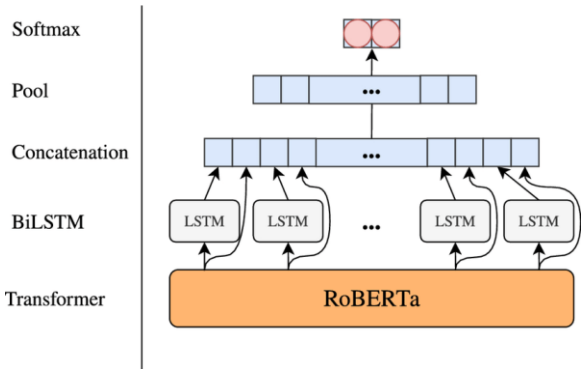


Fig. 1: System Overview

1. 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.

2. 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.

3. 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.

4. 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.).

5. 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.

6. 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.

7. 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.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental results demonstrated that the model effectively identified multiple emotional categories and sentiment classes from textual data with high classification accuracy. The contextual language understanding capability of RoBERTa enabled the model to capture semantic meaning, contextual dependencies, and subtle emotional expressions more effectively than traditional machine learning and conventional deep learning methods.

The analysis showed that the proposed approach maintained robust performance across different text domains, writing styles, and social media datasets while reducing misclassification of ambiguous expressions.

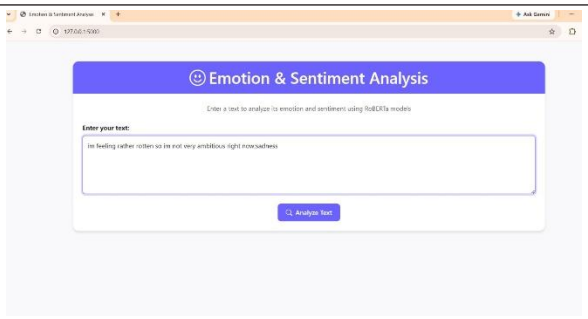


Fig. 2: Enter Text

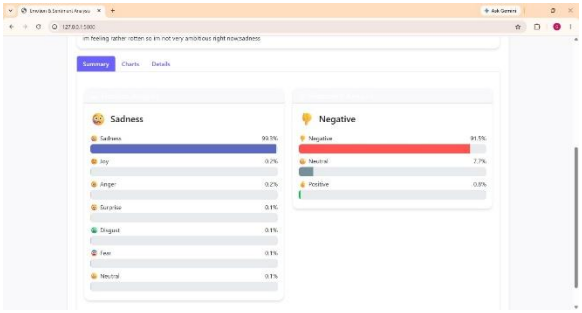


Fig. 3: Predicted Results

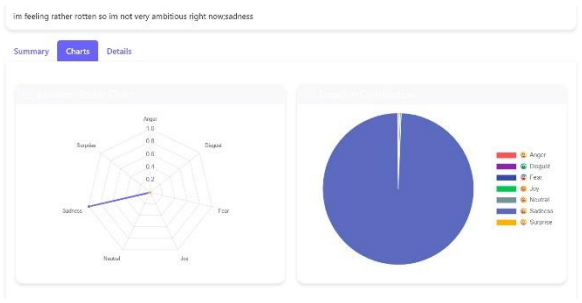


Fig. 4: Charts



Fig. 5: Graphs

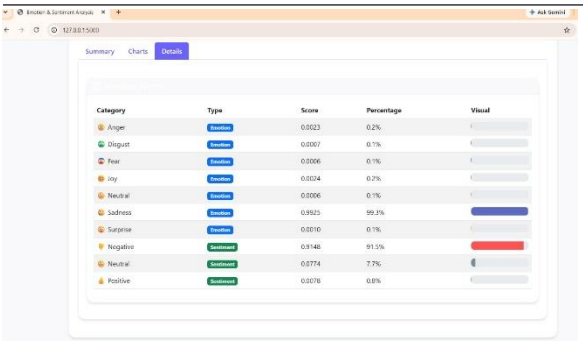


Fig. 6: Details

V. 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.

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. [Online]. Available: https://datareportal.com/reports/digital-2018-q3-global-digital-statshot

[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] Boyapati, P. K. Building a centralized data operations hub for healthcare enterprise integration. IJSAT-Int. J. Sci. Technol. 16 (2). https://doi.org/10.71097/IJSAT.v16.i2.3219

[8] Kandula, S. T. R., Susarla, R. S., & Boyapati, P. K. (2025, July). Enhanced Cyber Security Using Global Local Artificial Neural Network Based Intrusion Detection in Big Data Environment. In 2025 IEEE 4th World Conference on Applied Intelligence and Computing (AIC) (pp. 426-431). IEEE.

[9] Akinapalli, S. (2026). An Ai-Powered Data Trust And Quality Scoring Framework For Enterprise Decision Intelligence Systems. International Journal of Data Science and IoT Management System, 5(1), 946-950.

AUTHORS Profile



Mr. M. RAMAN KUMAR received his B.Tech degree in Computer Science and Engineering from Ranchi University in 2010 and his M.Tech degree in Computer Science and Engineering from Bhagwant University, Ajmer, in 2020. He is presently pursuing his Ph.D. from Atal Bihari Vajpayee Indian Institute of Information Technology and Management (ABV-IIITM), Gwalior. He has been dedicated to the teaching profession for the past 2 years. At present, he is working as an Assistant Professor in the Department of Computer Science and Engineering at Guru Nanak Institutions Technical Campus, Ibrahimpatnam, Telangana, India.



Ms. DONTHULA THRISHIKA pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Ms. GUNDEBOYINA ANUSHA pursuing B.Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. GUMPENA YASHWANTH pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.



Mr. ENIGA NITHINDHAR REDDY pursuing B. Tech degree in Computer Science and Engineering at Guru Nanak Institutions Technical Campus affiliated to JNTUH in 2022-2026.