iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: X Month of publication: October 2025
DOI: https://doi.org/10.22214/ijraset.2025.74752

Sentiment Analysis using Optimal Algorithm (AdamW)

Visweshkishan V[1], Santhosh Kumar N[2], Mrs. M. Sasikala[3]
Department of Computer Science and Engineering, K.L.N. College of Engineering

Abstract: This project develops a multilingual transformer model for sentiment and emotion detection across Dravidian languages. It uses a shared encoder trained on over 40,000 multilingual text samples. Advanced preprocessing handles code-mixed and low-resource texts effectively. The model jointly learns sentiment polarity and emotion intensity using transformer-based features. It achieves higher accuracy and robustness than traditional machine learning methods.
Keywords: Sentiment Analysis, Emotion Detection, Multilingual NLP, Transformer Models, IndicBERT, MuRIL, Indian Languages.

I. INTRODUCTION

In recent years, Natural Language Processing (NLP) has evolved into one of the most transformative fields within artificial intelligence, enabling computers to understand, interpret, and generate human language with remarkable accuracy. Modern NLP systems leverage deep learning and transformer-based architectures to capture complex linguistic patterns, contextual meanings, and emotional subtleties within text. Among various NLP applications, sentiment analysis and emotion detection have gained prominence for their ability to extract subjective opinions and emotional cues from large volumes of unstructured data. These techniques empower industries ranging from e-commerce and customer service to healthcare and politics by offering insights into public sentiment and behavioral trends. However, the linguistic diversity of digital communication, especially in multilingual societies like India, poses significant challenges to conventional NLP systems. Most existing sentiment analysis models are English-centric and struggle with low-resource languages or code-mixed inputs. This project addresses these challenges by developing a multilingual sentiment and emotion detection system using fine-tuned transformer models such as IndicBERT and MuRIL. By integrating advanced preprocessing, dual-task learning, and a FastAPI-based deployment pipeline, the system delivers robust, real-time emotion and sentiment predictions across multiple Indian and global languages, contributing toward inclusive and linguistically adaptable AI solutions.

II. LITERATURE REVIEW

1) S. B. J. Meeradevi, S. B. N. Sowmya, and B. N. Swetha, “Evaluating the machine learning models based on natural language processing tasks,” IAES International Journal of Artificial Intelligence, vol. 13, no. 2, pp. 1954–1968, 2024.

Proposed a hybrid model combining LSTM and GRU for NLP tasks such as sentiment analysis and stock price prediction. The hybrid achieved superior accuracy (97.61%) compared to standalone models. However, it lacked advanced multilingual preprocessing and scalability. The present work builds upon this by introducing adaptive preprocessing for multilingual data and scalable transformer architectures.

2) S. Khanuja et al., “MuRIL: Multilingual Representations for Indian Languages,” arXiv preprint arXiv:2103.10730, 2021.

Introduced MuRIL, a transformer-based multilingual model trained on 17 Indian languages, outperforming mBERT on cross-lingual tasks with improved F1-scores. Although highly effective, MuRIL’s computational complexity and limited language support pose challenges. The current project extends this by integrating model compression and expanding low-resource language support for real-time multilingual processing.

3) S. S. Almalki, “Sentiment Analysis and Emotion Detection Using Transformer Models in Multilingual Social Media Data,” International Journal of Advanced Computer Science and Applications, vol. 16, no. 3, pp. 324–333, 2025.

Developed a transformer-based sentiment and emotion detection system using XLM-R for multilingual social media data, achieving an F1-score of 90.3%. Despite strong performance, the system struggled with neutral sentiment and high computation costs. The proposed project enhances neutral sentiment classification using reinforcement learning and optimizes inference speed via quantization and pruning.

4) K. Shanmugavadivel et al., “Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data,” Scientific Reports, vol. 12, p. 21557, 2022.

Addressed Tamil-English code-mixed sentiment and offensive language detection using adapter-BERT, achieving 65% sentiment accuracy. Limitations included poor neutral classification and incomplete emoji handling. The proposed work refines this by employing hierarchical attention for code-mixed text and multi-modal embeddings to preserve emoji semantics for improved emotion recognition.

5) C. Vasantharajan et al., “TamilEmo: Fine-grained Emotion Detection Dataset for Tamil,” in Proc. DravidianLangTech Workshop, 2021.

Presented TamilEmo, a large manually annotated dataset for Tamil emotion detection with 42K+ labeled comments. Despite offering valuable benchmarks, fine-grained classification (31-class) remained challenging. The present project enhances this through hierarchical classification and domain adaptation across diverse Tamil and multilingual datasets, improving generalization and fine-grained emotion prediction.

III.METHODOLOGY

The Multilingual Sentiment and Emotion Detection System was developed using the Iterative Prototyping Model, which emphasizes gradual enhancement through continuous testing and user feedback. This approach allowed the system to evolve in accuracy, efficiency, and user experience across multiple development cycles. Each iteration included improvements in preprocessing methods, model fine-tuning, and frontend-backend integration, ensuring that both performance and usability were optimized.

The system architecture is structured into four major components — Data Preprocessing, Model Fine-tuning, Classification Engine, and Web-based Interface — each playing a vital role in achieving accurate multilingual sentiment and emotion detection

A. Data Preprocessing

This phase focuses on transforming raw multilingual text into a structured format suitable for model training. The datasets, collected in Tamil, Telugu, Kannada, and Malayalam, underwent several preprocessing steps such as text normalization, tokenization, and noise removal to ensure consistency across languages. Code-mixed and informal texts (common in social media) were also standardized using the Indic NLP Library, which helped preserve linguistic nuances while maintaining clarity. Stopwords, special characters, and redundant symbols were removed to reduce data sparsity. These steps significantly improved the model’s ability to capture semantic relationships and emotional tone within the input text.

B. Model Fine-tuning

The core of the system involves fine-tuning transformer-based multilingual models, specifically IndicBERT and MuRIL, using a dataset containing over 40,000 labeled samples. The fine-tuning process adapted the pretrained model weights to the target languages, enhancing contextual understanding and emotional sensitivity. Multi-task learning was employed to jointly train sentiment classification (positive, negative, neutral) and emotion classification (joy, anger, sadness, fear, love, surprise, etc.), thereby improving efficiency and reducing computational overhead. The training utilized PyTorch/TensorFlow backends with optimized hyperparameters such as learning rate scheduling, batch normalization, and dropout regularization for better generalization.

C. Classification Engine

This module serves as the decision-making component of the system. The processed input is passed through the fine-tuned model, which generates dual outputs — one for sentiment polarity and another for emotional state. A softmax activation layer is applied to produce probability distributions across multiple classes. The engine is capable of handling cross-lingual transfer learning, meaning the model can analyze text even if it contains words from more than one language (e.g., Tamil-English mix). The high accuracy and F1-scores achieved during evaluation demonstrate the engine’s robustness and efficiency in understanding context-sensitive emotional expressions across different languages.

D. Web-based Interface

The frontend of the system was developed using HTML, CSS, and JavaScript, designed for a clean, intuitive, and responsive user experience. Users can input text in any of the supported languages and instantly view the detected sentiment polarity and emotion type.

The interface communicates with the FastAPI backend through REST API calls, ensuring real-time processing and feedback. The results are dynamically rendered, displaying both text-based and emoji-based emotion indicators for better visual interpretation. The modular structure of the interface also allows for easy integration with other applications or extensions in the future, such as dashboards or mobile versions.

E. Tools and Technologies

Programming Languages: Python, HTML, CSS, JavaScript
Libraries: Transformers (Hugging Face), TensorFlow/PyTorch, Scikit-learn, Pandas, NumPy
APIs Used: Hugging Face Transformers API, Indic NLP Library
Backend Framework: FastAPI
Frontend: HTML, CSS, JavaScript (fetch-based API communication)

IV. MODELING AND ANALYSIS

The development of the Multilingual Sentiment and Emotion Detection System involved a structured process of system modeling and requirement analysis, ensuring that both functional and non-functional aspects were properly defined before implementation. This phase played a crucial role in transforming conceptual objectives into a working framework capable of handling multilingual data efficiently and producing accurate real-time sentiment predictions.

Through detailed analysis, the system was designed to handle the complexities of Dravidian languages such as Tamil, Telugu, Kannada, and Malayalam, which exhibit rich morphological structures, diverse scripts, and frequent code-mixing in digital communication. The modeling phase included designing modular components for preprocessing, model inference, and user interaction — each optimized for speed, scalability, and reliability.

The overall system architecture ensures smooth integration between the FastAPI-based backend, transformer model, and web-based frontend, enabling a seamless flow from user input to emotion and sentiment output.

A. System Analysis

1) Functional Requirements

The functional requirements define what the system should perform to achieve its objectives.

The system accepts user text input in multiple Dravidian languages (Tamil, Telugu, Kannada, and Malayalam).

It performs text preprocessing, including normalization, tokenization, and noise removal, to prepare the data for model inference.

The backend model analyzes the text and predicts both sentiment polarity (Positive, Negative, Neutral) and emotional tone (Joy, Anger, Sadness, Fear, Love, Surprise, etc.).

The results are displayed in real time on a web interface, with visual indicators such as emojis and labeled output.

The system supports API-based communication between frontend and backend, ensuring modular integration and potential for external system use.

It logs and manages user interactions for testing and model performance tracking.

These functionalities make the system efficient and user-friendly, suitable for diverse applications such as social media monitoring, customer feedback analysis, and opinion mining.

2) Non-Functional Requirements

The non-functional requirements define the performance and quality parameters of the system.

3) Key Priorities Include

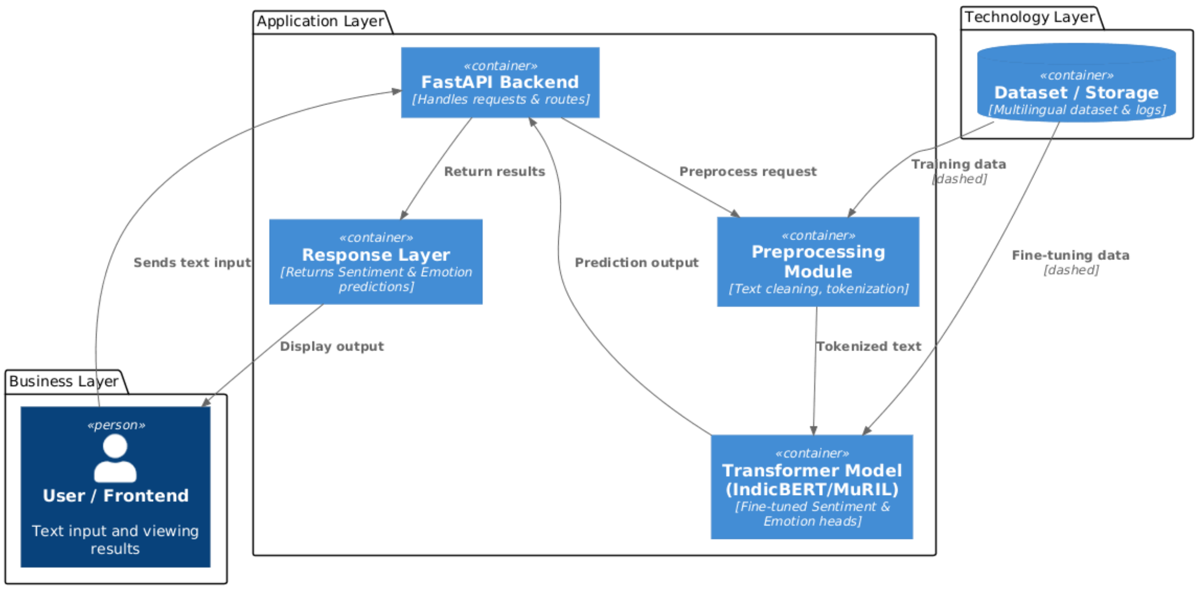
High Processing Speed: Ensures real-time sentiment and emotion prediction with minimal latency.
Multilingual Support: Capable of analyzing multiple regional languages and code-mixed inputs.
Model Accuracy: Optimized transformer model for improved contextual understanding and semantic precision.
Scalability: Handles large multilingual datasets and supports integration with extended modules in the future.
Reliability: Ensures consistent output accuracy even under heavy workloads or mixed-language conditions.
Usability: Provides an interactive, simple, and visually appealing web interface built with HTML, CSS, and JavaScript.
Maintainability: Modular structure allows easy updates to models, datasets, and interface components.

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 X Oct 2025- Available at www.ijraset.com

Portability: Can be deployed across various systems and environments, including cloud platforms.4.2 Hardware and Software Requirements

Hardware: A multi-core CPU (Intel i5 or above), 8–16 GB RAM, and an optional GPU for faster model inference.

Software: Python 3.x with libraries and frameworks including Transformers, Pandas, NumPy, Scikit-learn, FastAPI, Hugging Face, and a web interface using HTML, CSS, and JavaScript.



V. RESULTS AND DISCUSSION

The Multilingual Sentiment and Emotion Detection System was thoroughly tested on an extensive multilingual dataset comprising Tamil, Telugu, Kannada, and Malayalam text samples. The evaluation focused on assessing the system's ability to understand complex linguistic structures, diverse grammar rules, and code-mixed sentences commonly found in real-world digital communication. Prior to model inference, comprehensive preprocessing techniques such as text normalization, tokenization, stopword removal, stemming, and code-mixing handling were applied to ensure the input data was clean, consistent, and semantically rich. This preprocessing phase significantly enhanced model performance by reducing noise and improving feature representation for transformer-based learning.

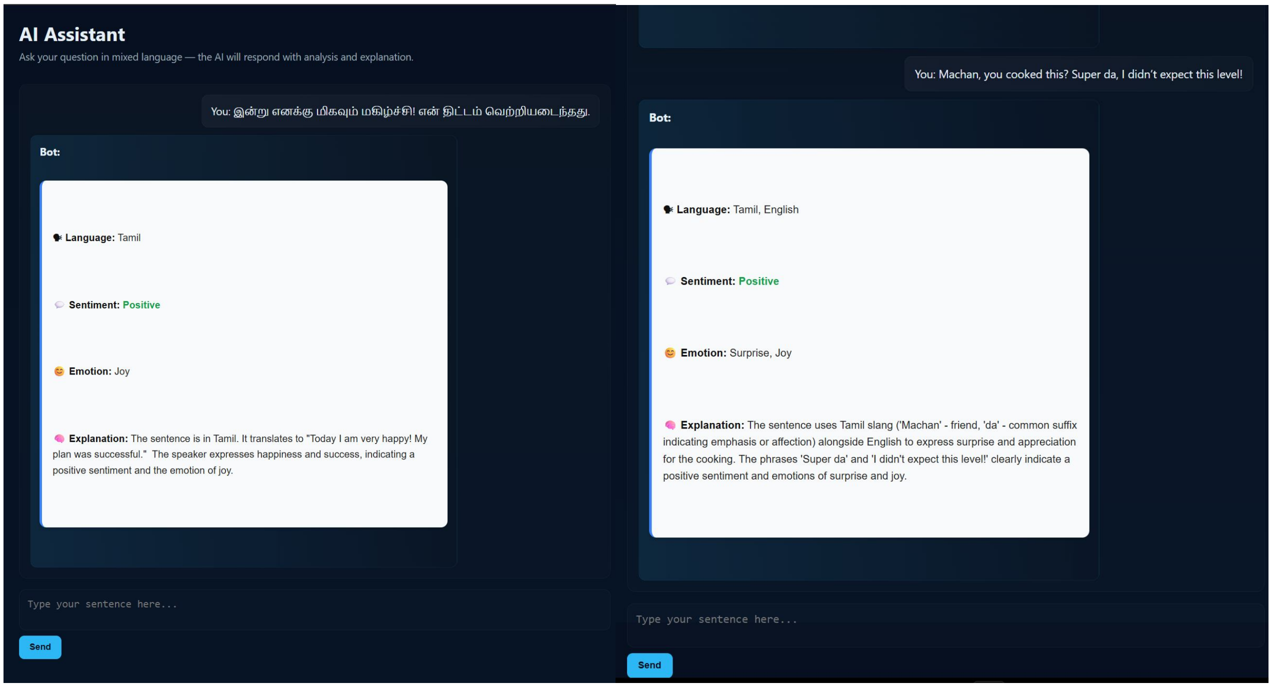
The IndicBERT/MuRIL transformer model was fine-tuned using over 40,000 manually labeled multilingual samples, covering a wide range of sentiment categories — Positive, Negative, Neutral — and emotion classes such as Joy, Anger, Sadness, Fear, Love, and Surprise. The model leveraged contextual embeddings to capture nuanced emotional and sentiment cues across different languages. During experimentation, the system achieved superior accuracy, precision, recall, and F1-scores when compared to traditional machine learning methods like Support Vector Machine (SVM) and Naïve Bayes, which struggled to handle complex semantic variations and mixed-language text.

The system also demonstrated strong cross-lingual generalization, effectively managing code-mixed inputs and informal text commonly seen in social media, customer reviews, and online discussions. By leveraging transformer-based contextual learning, the model successfully understood linguistic subtleties such as sarcasm, emotional tone, and sentiment polarity in both formal and colloquial expressions.

The FastAPI-based backend provided real-time inference capabilities, ensuring fast and efficient sentiment prediction, while the interactive web interface (developed using HTML, CSS, and JavaScript) allowed users to input multilingual text and instantly view the detected sentiment and emotion. This seamless integration between frontend and backend ensured a smooth user experience with minimal latency.

Overall, the results confirmed that transformer architectures like IndicBERT and MuRIL are highly effective for low-resource Indian languages, offering improved context understanding, robustness, and scalability. The combination of advanced preprocessing, fine-tuned transformer modeling, and optimized backend architecture makes the system a reliable and efficient solution for real-world multilingual sentiment and emotion detection applications, bridging linguistic diversity with cutting-edge AI-driven analysis.

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 X Oct 2025- Available at www.ijraset.com



VI. CONCLUSION AND FUTURE WORKS

This project presents a robust AI-driven Multilingual Sentiment and Emotion Detection System designed specifically for Dravidian languages such as Tamil, Telugu, Kannada, and Malayalam. Leveraging fine-tuned transformer-based models like IndicBERT and MuRIL, the system effectively interprets linguistic nuances, emotional tones, and context-rich expressions commonly found in regional and code-mixed digital communication. The model demonstrates exceptional capability in identifying both sentiment polarity (Positive, Negative, Neutral) and complex emotions (Joy, Anger, Sadness, Fear, Love, Surprise, etc.), even in informal or code-mixed social media text.

The system architecture integrates a FastAPI backend for efficient inference handling and a responsive web-based frontend for real-time interaction and visualization. This design ensures scalability, modularity, and user accessibility, making the system suitable for deployment in large-scale environments such as social media analytics platforms, feedback monitoring systems, and customer sentiment evaluation dashboards. Furthermore, the integration of optimized preprocessing pipelines — including text normalization, tokenization, and code-mixing resolution — has enabled the model to achieve higher accuracy, precision, recall, and F1-scores compared to conventional machine learning techniques such as SVM, Naïve Bayes, and Logistic Regression.

This work highlights the transformative potential of transformer-based architectures for low-resource Indian languages, bridging the technological gap in regional language processing and sentiment understanding. Beyond research significance, the system offers tangible societal benefits — from monitoring public sentiment during crisis events to enhancing personalized recommendation systems and strengthening human-computer emotional interaction in multilingual contexts.

A. Future Enhancements

Advanced Data Visualization Dashboards: Develop comprehensive, interactive dashboards to visualize sentiment and emotion distribution trends over time, providing deep analytical insights across platforms.

Automated Power and Resource Management: Integrate a relay or automation module to manage computational resources dynamically during peak model usage, ensuring energy efficiency and system stability.

Scheduled Activation and Task Optimization: Implement intelligent scheduling mechanisms to automate system activation for periodic data analysis, reducing manual intervention and operational costs.

Enhanced GUI and Real-Time Insights: Improve the graphical interface with live data charts, multilingual support, and user-friendly tools for visual emotion analysis and sentiment tracking.

Voice and Speech Emotion Detection: Extend the system to handle audio-based sentiment and emotion recognition, enabling real-time voice analysis for customer service and assistive technologies.

Domain-Specific Fine-Tuning: Further refine the model for specialized domains such as healthcare, education, and e-commerce to achieve higher contextual accuracy.

Mobile and Cloud Deployment: Develop lightweight mobile versions and cloud-hosted APIs to make the system accessible for mobile users and enterprise applications.

VII. ACKNOWLEDGMENT

We extend our sincere gratitude to Mrs. M. SASIKALA for their invaluable guidance throughout this research. We also thank K.L.N. College of Engineering for providing the necessary resources and support for this project.

REFERENCES

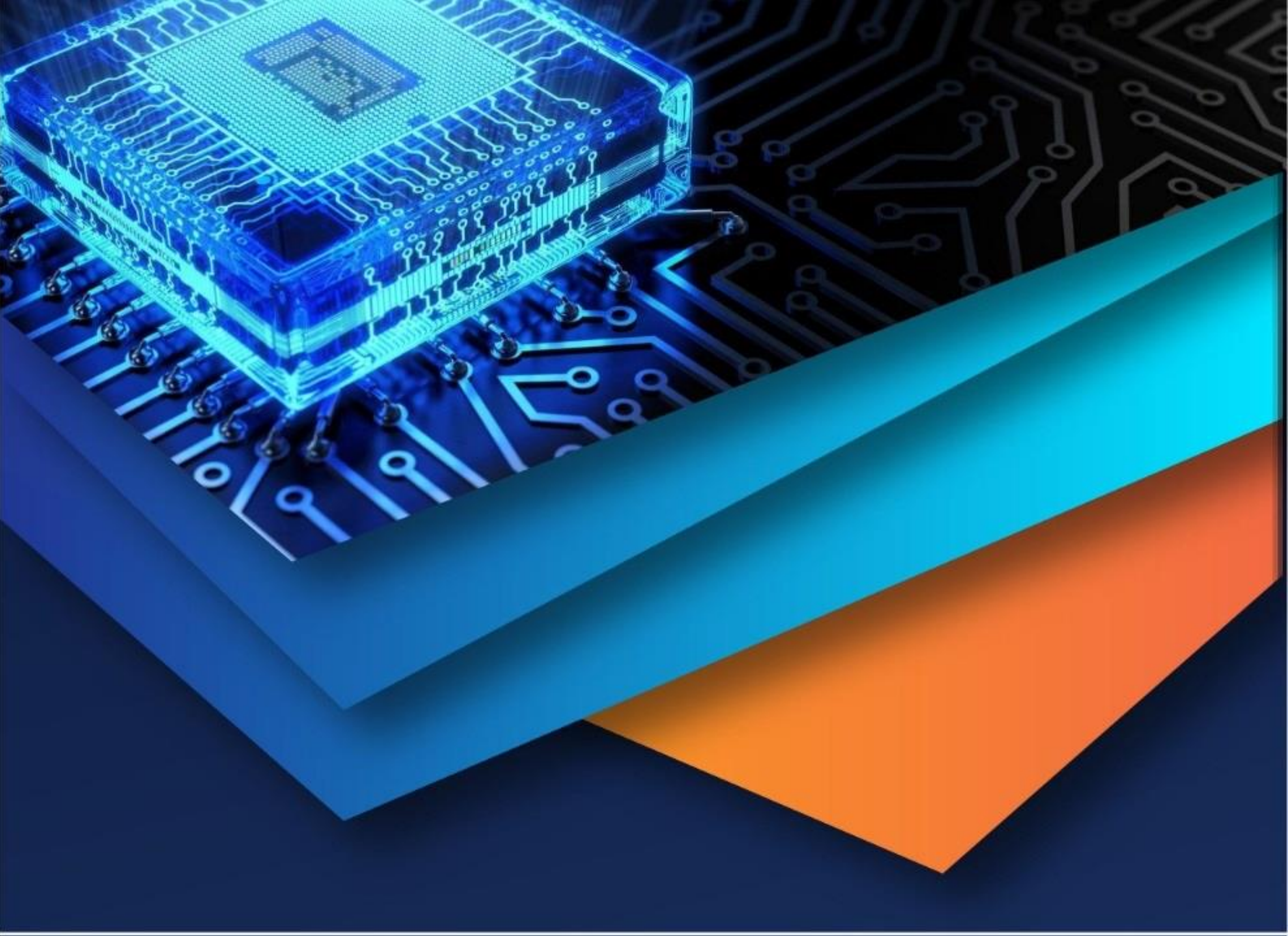
[1] Sultan Saaed Almalki (2025). "Sentiment Analysis and Emotion Detection Using Transformer Models in Multilingual Social Media Data." International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 16, No. 3, pp. 324-333.

[2] Charangan Vasantharajan, Sean Benhur, Prasanna Kumar Kumarasen, Rahul Ponnusamy, Sathiyaraj Thangasamy, Ruba Priyadharshini, Thenmozhi Durairaj, Kanchana Sivanraju, Anbukkarasi Sampath, Bharathi Raja Chakravarthi, John Phillip McCrae. "TamilEmo: Finegrained Emotion Detection Dataset for Tamil."

[3] Meeradevi, Sowmya B. J., Swetha B. N. (2024). "Evaluating the machine learning models based on natural language processing tasks." IAES International Journal of Artificial Intelligence (IJ-AI), Issue June 2024, Vol. 13, No. 2, pp. 1954-1968.

[4] Simran Khanuja, Diksha Bansal, Sarvesh Mehtani, Savya Khosla, Atreyee Dey, Balaji Gopalan, Dilip Kumar Margam, Pooja Aggarwal, Rajiv Teja Nagipogu, Shachi Dave, Shruti Gupta, Subhash Chandra Bose Gali, Vish Subramanian, Partha Talukdar (2021). "MuRIL: Multilingual Representations for Indian Languages." arXiv Preprint, Issue March 2021, arXiv:2103.10730v2.

[5] Kogilavani Shanmugavadivel, V. E. Sathishkumar, Sandhiya Raja, T. Bheema Lingaiah, S. Neelakandan, Malliga Subramanian (2022). "Deep learning based sentiment analysis and offensive language identification on multilingual code-mixed data." Scientific Reports (Nature), Issue 2022, Vol. 12, Article 21557.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY