iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: https://doi.org/10.22214/ijraset.2026.85312

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue X Oct 2026- Available at www.ijraset.com

CLARITAS: Digitizing Educational Feedback Collection Through AI-Driven Sentiment Analysis and Agentic Orchestration

Vishva Sruthi N K[1], Shahana R[2], Venkata Lakshmi S[3]
[1,2]*Student,* [3]*Associate Professor , Department of Computer Science and Engineering, K.L.N. College of Engineering, Sivagangai, Tamil Nadu*

Abstract: ***Traditional feedback collection in educational institutions is often paper-based, resulting in delayed analysis and limited utilization of student insights. This paper presents CLARITAS, an AI-powered educational feedback analytics platform designed to digitize and automate feedback collection and analysis for classrooms and training sessions. Featuring a role-based architecture for administrators, teachers, and students, the system analyses feedback using locally deployed DistilBERT for sentiment classification, Term Frequency-Inverse Document Frequency (TF-IDF) for keyword extraction, and word-frequency analysis. A self-built, restricted Agentic AI layer coordinates these components to generate structured summaries without relying on external LLMs or third-party APIs. A fully local analytical workflow supports data privacy, institutional governance, and traceability. The proposed approach provides a privacy-conscious and scalable alternative to conventional manual feedback analysis.***

Keywords: Sentiment Analysis, DistilBERT, TF-IDF, Agentic AI, Local AI, Educational Technology, Feedback Analytics, Natural Language Processing (NLP)

I. INTRODUCTION

Student feedback helps instructors and administrators understand learning experiences, teaching effectiveness, and areas needing improvement. However, feedback in most educational institutions is still collected through paper forms, spreadsheets, or basic online forms followed by manual review — an approach that becomes time-consuming and inconsistent as response volumes grow. Manual review is slow, struggles to identify common themes across unstructured comments, delays corrective action, and tends to underuse textual responses in favour of numerical ratings alone.

Recent advances in Natural Language Processing (NLP) make it possible to automate this analysis: sentiment analysis reveals the emotional orientation of responses, while keyword extraction and word-frequency analysis surface recurring topics. Transformer-based models such as DistilBERT enable efficient contextual sentiment classification. However, individual NLP techniques alone do not constitute a complete analytics solution — a practical system must integrate collection, analysis, and interpretation into one workflow while addressing privacy and usability.

Privacy is a particular concern when feedback is processed using external AI services, since student responses may contain sensitive opinions and transferring them externally raises data-governance and transparency issues. This motivates an approach that performs AI-assisted analysis locally, coordinates multiple analytical techniques, and generates structured insights without transmitting feedback to external AI services.

This paper proposes CLARITAS, an AI-powered educational feedback analytics platform with role-based access for administrators, teachers, and students. Textual responses are analysed using a locally deployed DistilBERT model for sentiment classification, combined with TF-IDF-based keyword extraction and word-frequency analysis. These outputs are coordinated by a self-built, restricted agentic AI orchestration layer that synthesizes them into structured feedback summaries through a controlled workflow, without relying on any external large language model or third-party AI API — keeping the entire analysis pipeline local.

The major contributions of this work are as follows:

1) An integrated educational feedback analytics platform that supports structured feedback collection and automated analysis through role-based access for administrators, teachers, and students.

2) A multi-component NLP analysis pipeline combining locally deployed DistilBERT-based sentiment classification, TF-IDF-based keyword extraction, and word-frequency analysis to derive complementary insights from textual feedback.

©IJRASET: All Rights are Reserved

3) A self-built, restricted Agentic AI orchestration layer that coordinates these tools and generates structured feedback summaries through a rule-based workflow without relying on external large language models or third-party AI APIs.
4) A privacy-preserving local processing workflow that keeps feedback analysis within the local deployment environment without transmitting data to external AI services.
5) An analytics dashboard that presents sentiment distributions, frequently occurring terms, keyword insights, and structured summaries for efficient interpretation.

II. RELATED WORK

Recent research in sentiment analysis has increasingly adopted transformer-based language models, particularly BERT and its lightweight variants, to improve contextual understanding while reducing computational requirements. The following studies demonstrate the application of transformer models and complementary text-analysis techniques in different domains.

Joshy and Sundar [1] compared BERT, DistilBERT, and RoBERTa for sentiment analysis using movie reviews and tweets. Their study demonstrated the effectiveness of DistilBERT for sentiment classification with reduced computational requirements. This supports the use of DistilBERT in CLARITAS for efficient classification of student feedback.

Amanullah et al. [6] investigated DistilBERT for sentiment analysis of user-generated reviews. Their work demonstrated the applicability of DistilBERT in identifying sentiment from textual feedback. This is relevant to CLARITAS, where student responses are classified into Positive, Neutral, and Negative categories.

Özmen and Yıldırım [5] explored a hybrid approach combining DistilBERT and TF-IDF for text analysis. Their study highlighted the complementary nature of contextual representations and statistical text features. Based on this approach, CLARITAS employs DistilBERT for sentiment classification and TF-IDF for extracting significant keywords from student feedback.

However, existing studies primarily focus on individual sentiment-analysis or text-analysis tasks. An integrated approach that combines sentiment classification, keyword extraction, visualization, rating analysis, and automated summarization for educational feedback is comparatively limited. CLARITAS addresses this gap by providing a unified platform for analysing student feedback and generating meaningful insights.

III. PROPOSED SYSTEM

CLARITAS automates the collection, processing, and interpretation of educational feedback, addressing the limitations of manual review and delayed insights. It integrates role-based access, structured feedback collection, NLP-based analysis, and a restricted Agentic AI orchestration layer into a unified workflow.

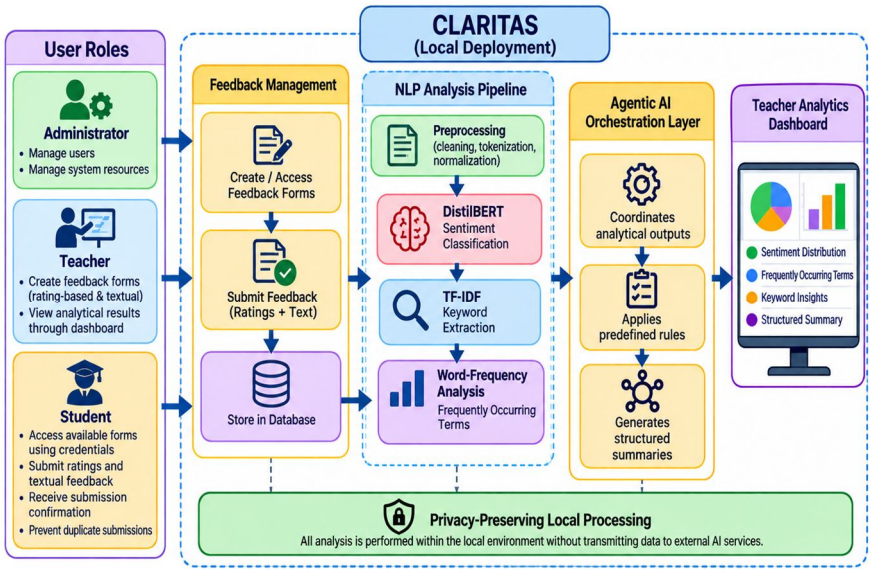


Fig.1 Overall System Architecture of CLARITAS

As shown in Fig. 1, the system combines role-based authentication, feedback management, database storage, NLP analysis, and the Agentic AI orchestration layer. Administrators manage users and system resources, teachers create feedback forms and view analytical results, and students access available forms using their credentials. Students submit rating-based and textual feedback, receive confirmation after successful submission, and are prevented from submitting multiple responses to the same form. Submitted feedback is stored in the database for automated analysis. The textual feedback is analysed using three NLP components: a locally deployed DistilBERT model for sentiment classification, TF-IDF for keyword extraction, and word-frequency analysis for identifying frequently occurring terms. These components provide complementary insights into sentiment, important keywords, and recurring topics.

A self-built, restricted Agentic AI orchestration layer coordinates the outputs of these analytical components through a predefined, rule-based workflow. It organizes and synthesizes the analytical results into structured feedback summaries without relying on external large language models or third-party AI APIs. The complete analytical workflow, including sentiment classification, keyword extraction, word-frequency analysis, orchestration, and summary generation, operates within the local deployment environment. Student feedback is not transmitted to external AI services, supporting data privacy and reducing external dependencies. The resulting insights are presented through the teacher analytics dashboard for efficient interpretation of student feedback.

IV. METHODOLOGY

The methodology of CLARITAS automates feedback collection, processing, analysis, and interpretation through a privacy-conscious workflow combining role-based feedback management, multiple NLP techniques, and a self-built restricted Agentic AI orchestration layer, with all analysis performed locally.

A. Feedback Data Collection

Teachers create structured feedback forms with rating-based and text-based questions, which are made available to students. Students authenticate using their roll numbers and submit responses, which are stored in MySQL and linked to the corresponding form. A duplicate-submission mechanism prevents repeat responses from the same student.

B. Feedback Preprocessing

Submitted textual feedback passes through a preprocessing stage before being forwarded to the sentiment classification, keyword extraction, and word-frequency components.

C. Sentiment Classification Using DistilBERT

A locally deployed DistilBERT model classifies the sentiment of each textual response, producing an overall distribution of positive, negative, and neutral feedback without requiring transmission to an external AI service.

D. TF-IDF-Based Keyword Extraction

TF-IDF identifies important terms based on their frequency within a document relative to the full collection:

$$TFIDF(t, d) = TF(t, d) \times IDF(t)$$

$$IDF(t) = log(N / DF(t))$$

where TF(t,d) is the frequency of the term t in document d, N is the total number of documents, and DF(t) is the number of documents containing t. The resulting scores identify significant keywords reflecting topics discussed by students.

E. Word-Frequency Analysis

This technique identifies terms occurring repeatedly across responses, complementing TF-IDF by highlighting commonly discussed topics, concerns, or positive aspects.

F. Analytical Output Integration

Sentiment, keyword, and frequency outputs are combined rather than treated independently and are coordinated through the restricted Agentic AI orchestration layer before being presented to teachers.

G. *Restricted Agentic AI Orchestration*

The orchestration layer coordinates outputs from DistilBERT, TF-IDF, and word-frequency analysis, validating and combining them using predefined rules to produce a structured feedback representation. It does not rely on external LLMs or third-party APIs, improving traceability and control over the analytical workflow.

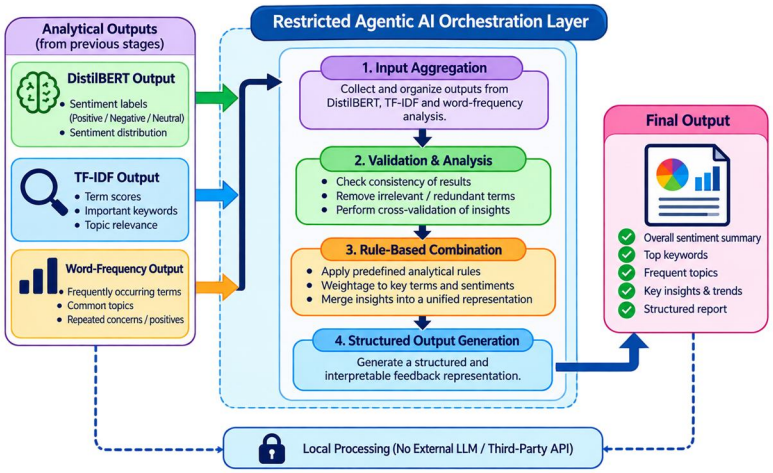


Fig.2 Restricted Agentic AI Orchestration Workflow

The restricted Agentic AI orchestration layer coordinates the analytical outputs generated by DistilBERT, TF-IDF, and word-frequency analysis. As shown in Fig. 2, these outputs are aggregated, validated, and combined using predefined rules to generate a structured feedback representation. The complete orchestration process operates locally without relying on external LLMs or third-party AI APIs.

V. SYSTEM ARCHITECTURE AND DESIGN

A. *System Architecture*

CLARITAS follows a layered web-based architecture comprising the frontend, backend, database, NLP and AI, and analytics layers. The system provides role-based functionalities for administrators, teachers, and students. The frontend manages user interaction, while the FastAPI backend handles authentication, feedback operations, database communication, and analytical processing. MySQL provides persistent storage for users, feedback forms, questions, and responses. The NLP layer uses locally deployed DistilBERT, TF-IDF, and word-frequency analysis, while the restricted Agentic AI orchestration layer coordinates the analytical outputs to generate structured feedback insights. The results are presented through the teacher analytics dashboard, ensuring modularity and maintainability.

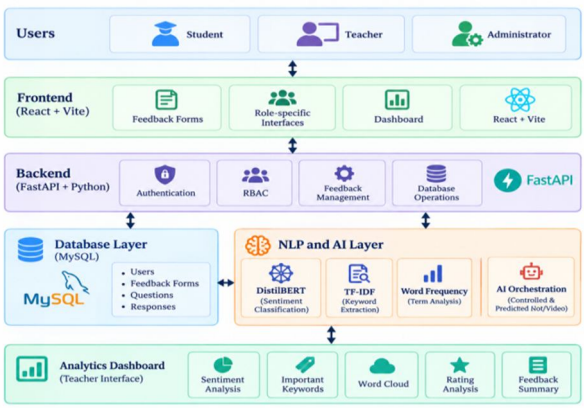


Fig.3 Layered Architecture and Component Interaction of CLARITAS

B. Frontend Design

The CLARITAS frontend is developed using React with Vite. It provides role-specific interfaces for students, teachers, and administrators. Students can access feedback forms, submit ratings and textual responses, while teachers can create forms and view analytics. Administrators manage users and system resources. Axios facilitates communication between the frontend and backend through REST APIs. React's component-based structure supports reusable and maintainable interfaces.

C. Backend Design

The backend is developed using FastAPI and Python and serves as the central application layer. It manages authentication, role-based access, feedback forms, submissions, database operations, and analytical processing. The backend coordinates the NLP workflow by passing textual feedback to DistilBERT, TF-IDF, and word-frequency components, after which the outputs are processed by the restricted Agentic AI orchestration layer and returned to the dashboard.

D. Database Design

MySQL is used for persistent storage of users, feedback forms, questions, and student responses. Relationships between students, forms, and submissions support efficient feedback retrieval and processing. The database also supports duplicate-submission prevention, ensuring that a student submits only one response for a given feedback form. Database operations are managed through the FastAPI backend according to user roles.

E. Authentication and Role-Based Access Control

CLARITAS implements JWT-based authentication and role-based access control for administrators, teachers, and students. Administrators manage system resources, teachers create forms and access analytics, and students submit feedback. Protected backend endpoints verify authentication and user roles before granting access to restricted functionalities.

F. NLP and AI Module Design

The NLP module processes textual feedback using three primary components: DistilBERT for sentiment classification, TF-IDF for keyword extraction, and word-frequency analysis for identifying frequently occurring terms. Their outputs are passed to the restricted Agentic AI orchestration layer, which follows a predefined workflow to organize and synthesize the results into structured feedback insights. The complete analysis workflow is locally deployed and does not require external LLM or third-party AI APIs.

G. Analytics Dashboard Design

The teacher analytics dashboard presents the processed feedback results through sentiment patterns, important keywords, frequently occurring terms, visualizations, and structured summaries. By consolidating these outputs into a single interface, the dashboard reduces manual examination of individual responses and helps teachers identify recurring patterns and areas requiring attention.

VI. IMPLEMENTATION

A. Development Environment

CLARITAS is implemented as a client-server web application using React/Vite for the frontend, FastAPI/Python for the backend, and MySQL for persistent storage. Axios manages frontend-backend communication, while JWT-based authentication provides secure role-based access. The environment integrates feedback management, NLP processing, AI orchestration, and analytics into a unified application.

B. Frontend Implementation

The frontend is developed using React and Vite and provides role-specific interfaces for students, teachers, and administrators. Students access forms and submit rating-based and textual feedback, while teachers create forms and view analytics. Administrators manage users and system resources. Axios handles communication with the backend.

C. Backend Implementation

The FastAPI/Python backend manages authentication, role verification, feedback operations, database communication, and analytical processing. Submitted responses are stored in MySQL, while textual feedback is routed to the NLP components and the resulting outputs are coordinated through the restricted Agentic AI orchestration layer.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Database Implementation

MySQL stores users, feedback forms, questions, and submitted responses, with relationships linking responses to their corresponding forms. Database operations are handled through the backend, which also implements duplicate-submission prevention for student-form combinations.

E. Authentication and Role Management

CLARITAS uses JWT-based authentication and role-based access control. Students authenticate using their roll number, teachers using their Staff ID, and administrators using their assigned credentials. Backend authorization restricts each user to the functionalities associated with their assigned role.

F. Feedback Form Implementation

Teachers create feedback forms containing rating-based and textual questions, while students submit their responses through the corresponding forms. Responses are stored in MySQL and linked to the relevant form. Duplicate-submission prevention ensures that each student can submit only once per form.

G. NLP Implementation

The NLP module uses a locally deployed DistilBERT model for sentiment classification, TF-IDF for keyword extraction, and word-frequency analysis for identifying recurring terms. The complete NLP workflow operates locally without external LLM or third-party AI API dependencies.

H. Agentic AI Orchestration Implementation

The self-built, restricted Agentic AI orchestration layer coordinates the outputs of DistilBERT, TF-IDF, and word-frequency analysis through predefined rules to generate a structured feedback summary. The workflow operates within fixed analytical boundaries without external LLM or third-party API dependency.

I. Analytics Dashboard Implementation

The analytics dashboard presents sentiment results, important keywords, frequent terms, visualizations, and structured summaries in a unified interface. This enables teachers to identify recurring feedback patterns and areas requiring attention without manually examining every response.

VII. RESULTS AND DISCUSSION

The implemented CLARITAS system successfully performs role-based authentication, feedback collection, automated feedback analysis, and visualization of analytical results. Students can submit rating-based and textual feedback, while teachers can create feedback forms and access the generated analytical insights.



Fig.4 Rating Distribution of Student Feedback.

The rating distribution shown in Fig. 4 represents the number of student responses received for each rating from 1 to 5 stars. The visualization provides a quantitative view of student feedback and helps teachers identify the overall pattern of ratings across the collected responses.



Fig.5 Sentiment Distribution of Student Feedback.

The sentiment distribution shown in Fig. 5 classifies the textual feedback into positive, neutral, and negative categories using the locally deployed DistilBERT model. This provides teachers with an overall understanding of the sentiment expressed by students.

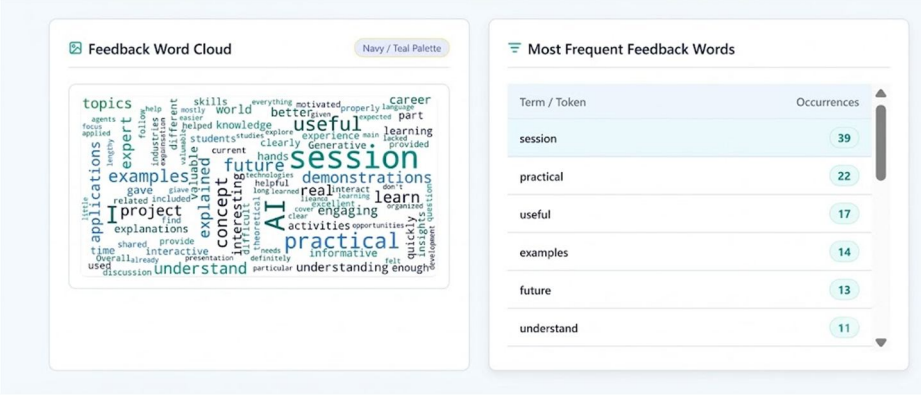


Fig. 6 Textual Feedback Analysis Using Word Cloud and Frequent Words

As shown in Fig. 6, the word cloud highlights prominent terms appearing in the textual responses, while the frequent-word analysis provides the corresponding occurrence counts. These outputs help identify recurring topics and commonly discussed aspects of the learning experience.

The outputs from sentiment classification, TF-IDF keyword extraction, and word-frequency analysis are then coordinated through the restricted Agentic AI orchestration layer to generate structured feedback summaries using predefined rules. The resulting insights are presented through the teacher analytics dashboard, enabling efficient interpretation of both rating-based and textual feedback. Overall, the results demonstrate that CLARITAS provides an integrated approach for automated educational feedback analysis. By combining quantitative ratings with NLP-based textual analysis, the system reduces manual analysis effort and enables teachers to identify important feedback patterns and recurring topics more efficiently. The local processing approach also ensures that student feedback is not transmitted to external AI services.

VIII. CONCLUSION

CLARITAS presents an AI-powered educational feedback analytics system that automates feedback collection, analysis, and visualization through a unified platform. The system integrates locally deployed DistilBERT for sentiment classification, TF-IDF for keyword extraction, and word-frequency analysis for identifying meaningful patterns in textual feedback. A restricted Agentic AI orchestration layer coordinates these analytical components through a controlled workflow to generate structured feedback summaries without relying on external LLMs or third-party AI APIs. By performing the analysis locally, CLARITAS helps reduce the need to transmit student feedback to external AI services, supporting a privacy-conscious approach to educational data processing. Overall, the system provides an integrated approach for transforming student feedback into structured insights that can assist educators in identifying trends and areas requiring attention.

©IJRASET: All Rights are Reserved

REFERENCES

[1] A. Joshy and S. Sundar, "Analyzing the Performance of Sentiment Analysis using BERT, DistilBERT, and RoBERTa," in Proc. IEEE Int. Power Renew. Energy Conf. (IPRECON), Kollam, India, 2022, pp. 1–6, doi: 10.1109/IPRECON55716.2022.10059542.

[2] H. Xiao and L. Luo, "An Automatic Sentiment Analysis Method for Short Texts Based on Transformer-BERT Hybrid Model," IEEE Access, vol. 12, pp. 93305–93317, 2024, doi: 10.1109/ACCESS.2024.3422268.

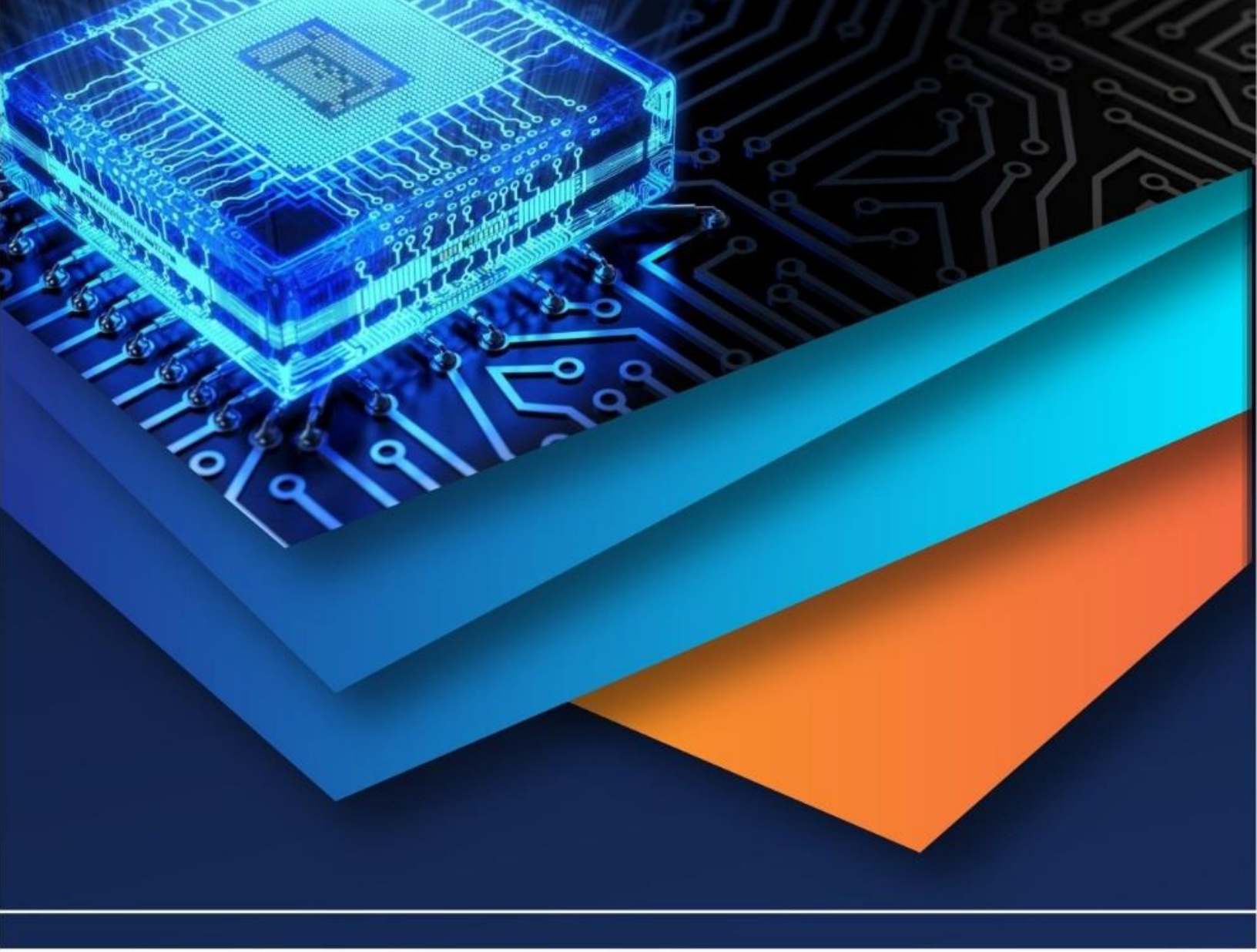
[3] C.-S. Lin, C.-N. Tsai, J.-S. Jwo, C.-H. Lee, and X. Wang, "Heterogeneous Student Knowledge Distillation From BERT Using a Lightweight Ensemble Framework," IEEE Access, vol. 12, pp. 33079–33088, 2024, doi: 10.1109/ACCESS.2024.3372568.

[4] P. Riedel, M. Reichert, R. von Schwerin, A. Hafner, D. Schaudt, and G. Singh, "Performance Analysis of Federated Learning Algorithms for Multilingual Protest News Detection Using Pre-Trained DistilBERT and BERT," IEEE Access, vol. 11, pp. 134009–134022, 2023.

[5] Ü. Özmen and E. O. Yıldırım, "DistilBERT-Based Hybrid Architecture for Phishing URL Detection," IEEE Access, vol. 14, pp. 71720–71737, 2026, doi: 10.1109/ACCESS.2026.3684855.

[6] Amanullah, S. Ullah, R. T. Chohan, and M. N. Khan, "Understanding Perceived Quality of Mobile Apps Using DistilBERT Model," IEEE Access, vol. 14, pp. 3560–3576, 2026, doi: 10.1109/ACCESS.2025.3648279.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)