iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84161

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 VII July 2026- Available at www.ijraset.com

Smart Complaint Detection and Severity Analysis in Financial Texts

Shravya Koulammagari[1], B. Kranthi Kiran[2]
[1]M. Tech Student Department of Computer Science and Engineering, JNTU Hyderabad, Telangana
[2]Professor of Computer Science and Engineering, JNTU Hyderabad, Telangana

Abstract: Financial institutions receive thousands of complaints daily across digital platforms such as social media, email, and customer service portals. Manual identification, categorization, and resolution of these complaints are slow, inefficient, and error-prone, leading to delayed responses, customer frustration, and missed opportunities for service improvement. The overwhelming volume of unstructured textual data makes it nearly impossible for human agents to process, prioritize, and address every complaint in a timely and effective manner. To address this challenge, this study presents a Multi-Task Financial Complaint Analysis System that leverages a shared RoBERTa architecture to simultaneously perform four classification tasks: aspect classification, severity prediction, emotion detection, and sentiment analysis. The proposed model integrates a pre-trained roberta-base network as the shared encoder with task-specific attention mechanisms and classifier heads for each task. Focal loss and weighted sampling are incorporated to handle class imbalance and enhance model robustness. The framework is trained and evaluated on the Consumer Financial Protection Bureau (CFPB) Consumer Complaint Database containing real-world financial complaints and compared with several baseline approaches. Experimental results demonstrate that the proposed Multi-Task RoBERTa model achieves an overall accuracy of 91.88% and an F1 score of 88.28%, with per-task performance of 76% for aspect classification, 92% for severity prediction, 100% for emotion detection, and 96% for sentiment analysis. For practical deployment, the system is implemented as a Streamlit-based web application capable of real-time complaint analysis with confidence scores and probability distribution visualizations. The proposed framework offers an accurate, interpretable, and reliable solution for automated complaint understanding and prioritization in financial customer service environments.
Keywords: Multi-Task Learning, Financial Complaint Analysis, RoBERTa, Aspect Classification, Severity Prediction, Emotion Detection, Sentiment Analysis, Natural Language Processing, CFPB Dataset, Streamlit, Focal Loss.

I. INTRODUCTION

Financial institutions receive thousands of complaints daily across digital platforms such as social media, email, and dedicated customer service portals. These complaints contain valuable information about customer dissatisfaction, product issues, and service failures. However, the manual identification, categorization, and resolution of these complaints remain slow, inefficient, and error-prone processes that lead to delayed responses, customer frustration, and missed opportunities for service improvement. The overwhelming volume of unstructured textual data makes it nearly impossible for human agents to process, prioritize, and address every complaint in a timely and effective manner. The lack of automated systems capable of handling this scale of data creates significant operational inefficiencies and diminishes customer trust and satisfaction. Furthermore, existing automated solutions are often limited in scope, focusing on a single task such as sentiment analysis or basic classification, without addressing the multi-faceted nature of complaints that contain information about the product involved, the urgency of the issue, the customer's emotional state, and the overall sentiment expressed.

Recent advances in artificial intelligence and natural language processing have opened new possibilities for automating the analysis of customer feedback. The emergence of powerful transformer-based language models, such as RoBERTa, has enabled machines to understand the nuances of human language with remarkable accuracy. These models can be fine-tuned to perform specific tasks like text classification, sentiment analysis, and emotion detection, making them ideal for processing and understanding financial complaints at scale. Additionally, multi-task learning has emerged as a powerful paradigm enabling models to leverage shared representations across multiple related tasks, improving generalization and performance. By training on several objectives simultaneously, models can recognize commonalities and task-specific distinctions, allowing related tasks to reinforce one another. This approach is particularly beneficial for complaint analysis where tasks such as aspect classification, severity prediction, emotion detection, and sentiment analysis are inherently related and can benefit from shared knowledge.

This project introduces the design and implementation of a Multi-Task Financial Complaint Analysis System, an intelligent platform built to automatically understand, categorize, and prioritize customer complaints in the financial domain. Leveraging the power of advanced transformer models and multi-task learning, the system simultaneously predicts four critical aspects of each complaint: the product or service involved, the urgency level, the customer's emotional state, and the overall sentiment expressed in the text. The system is built on a strong technological stack comprising PyTorch for deep learning, Hugging Face Transformers for the RoBERTa model, and Streamlit for a responsive and accessible front-end interface. The system is trained on the CFPB Consumer Complaint Database, which contains real-world complaints about financial products and services, ensuring practical relevance and generalizability. The interactive Streamlit dashboard enables users to input complaint text, receive instant predictions with confidence scores, and visualize results, making the system practical for real-world deployment in financial customer service environments. Through its comprehensive multi-task approach, the system aims to bridge the gap between cutting-edge AI research and practical, real-world applications in financial customer service and complaint management.

II. RELATED WORK

Transformer-based language models have become a powerful tool for text classification and natural language understanding, with recent significant progress in financial complaint analysis [1]. Singh, Bhatia, and Saha proposed a multi-task framework using RoBERTa that classifies complaints and their associated attributes, highlighting the importance of integrating emotion and sentiment analysis in complaint detection [2]. Joshi, Kumar, and Patel introduced a multitasking model applied to financial tweets from bank customers, classifying complaint label, severity label, emotion, and sentiment simultaneously [3]. Jain et al. developed a bilingual financial complaint classification framework using transformer-based architectures, demonstrating the effectiveness of these models for multilingual financial text analysis [4]. While various methods have been proposed for complaint analysis, it is critical to design a framework that can incorporate real-time interpretability, consistency in prediction, and comprehensive understanding of complaint text so as to overcome the remaining challenges in this domain [5].

The financial complaint analysis task requires the ability to capture and interpret subtle variations in customer grievances as reflected in their textual descriptions [6]. In general, such variations manifest in the form of different complaint types, varying urgency levels, and diverse emotional expressions [7]. A number of techniques have been proposed for capturing the linguistic patterns producing complaints, including variations in word choice, sentence structure, and contextual meaning [8]. Owing to heterogeneity of complaints across different financial products, services, and customer demographics, developments in this area have involved various aspects, including representation of textual features, temporal modeling of complaint patterns, the development of complaint taxonomies, and classification techniques [9]. Recent research has also investigated advanced textual representation and contextual understanding techniques to enhance complaint analysis performance [10]. Das et al. explored aspect-based complaint identification using video complaint datasets, demonstrating improved structural understanding of complaints across multiple modalities [11].

Human perceptions of intelligent systems increasingly rely on transparency and interpretability [12]. Wang et al. proposed a hybrid machine learning framework integrating complaint narratives with topic representations and structured attributes for predicting complaint outcomes [13]. Similarly, Das et al. introduced an interpretability-focused framework for distinguishing negative reviews from actual complaints, providing visual explanations of model predictions [14]. Both methods greatly promoted the explainable artificial intelligence field in financial applications [15]. However, most financial complaint analysis systems still lack integrated interpretability, which is essential for real-time deployment where the workloads of interpretability can be reduced and the risk of misbehavior can be minimized.

III. METHODOLOGY

A. System Overview

The proposed Multi-Task Financial Complaint Analysis System is designed to perform real-time complaint understanding by combining deep learning and multi-task learning techniques. Using the CFPB Consumer Complaint Database, the system extracts semantic features from complaint text through a shared RoBERTa encoder and captures task-specific patterns using dedicated attention mechanisms and classifier heads. To improve classification accuracy and feature learning, a hybrid loss function incorporating focal loss and weighted task losses was employed. The framework was further deployed as a Streamlit-based web application supporting real-time complaint analysis, while confidence scores and probability distributions were integrated to provide explainable and interpretable predictions.

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 VII July 2026- Available at www.ijraset.com

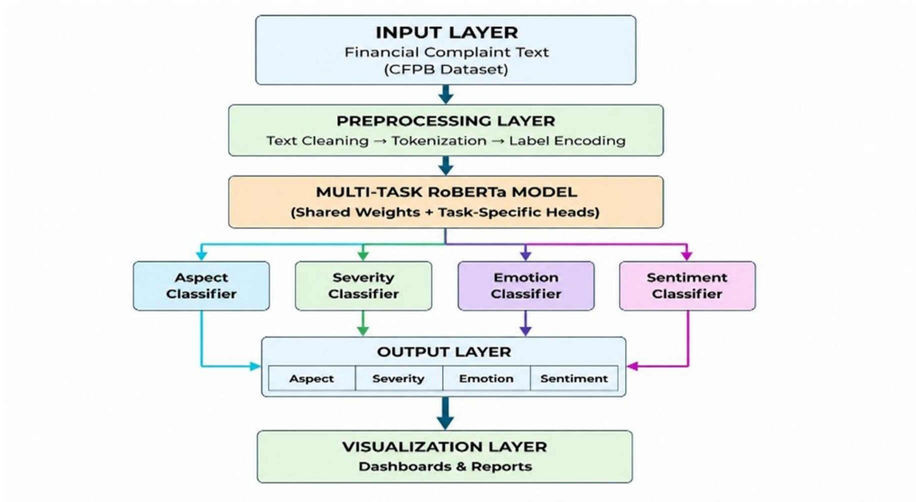


Fig.1 System Architecture

Fig. 1 shows the architecture of the proposed real-time financial complaint analysis framework, which contains the deployment framework, the feature learning and multi-task classification module with RoBERTa and task-specific heads, the classification module, and the visualization module with confidence scores and probability distributions. The proposed framework analyzes the semantic and contextual representations of complaints jointly. It can be deployed with Streamlit and run in real time, providing a flexible platform for intelligent complaint analysis applications through web-based interfaces for real-time complaint predictions. Confidence scores and probability distributions are used to identify the most influential textual features that cause the model to classify complaints into specific categories.

B. Data Acquisition

The financial complaint analysis framework has been implemented and tested on the CFPB Consumer Complaint Database, which is publicly available from the Consumer Financial Protection Bureau. The CFPB dataset consists of consumer complaints about financial products and services, containing information such as the product involved, the issue, the company, and the consumer's narrative description. The dataset includes complaints with variations in text length, complaint formats, and expression intensity. Appropriate training, validation, and testing splits ensure that the complaint analysis framework will generalize well for real-time usage.

C. Preprocessing

Preprocessing processes were applied to increase data performance, augment learning consistency, reduce class imbalances, and improve textual representations, thereby facilitating accurate, reliable, and robust real-time complaint analysis performance.

1) *Dataset Preparation and Organization:* The CFPB Consumer Complaint Database was categorized into different complaint categories based on product type and issue. The text and label information were organized into data collections so that the dataset could be prepared for subsequent training, validation, and testing procedures. The categorization of the dataset made it easy for data handling and improved the consistency between training and testing sets.
2) *Train–Validation–Test Data Partitioning:* The complete dataset was separated into training, validation, and testing subsets to improve model generalization and prevent overfitting. Stratified partitioning was performed accordingly. The training, validation, and testing subsets contained 80%, 10%, and 10% of the samples, respectively. Appropriate dataset partitioning further improves the reliability of analysis results, prevention of overfitting, and success of testing the complaint understanding ability on realistic complaint data
3) *Class Imbalance Mitigation:* During model training, the imbalanced distribution among complaint categories was countered with an adaptive weighted sampling approach so that minority classes receive a proportionally higher weighting. The balancing strategy particularly contributed to the improvement of the classification performance on the under-represented classes, such as specific product categories and emotions. Weighted random sampling was employed to counterbalance the distribution among complaint categories, promoting fair training and improving overall recognition accuracy.

4) *Text Preprocessing and Standardization:* Complaint text was preprocessed for feature learning including lowercasing, removing URLs, special characters, and extra whitespace to acquire consistent input representation. Text normalization allowed numerical consistency during feature learning while standardizing all training samples to a specific format. Uniform preprocessing allowed for consistent text representation which improved convergence during training and allowed for the efficient extraction of complaint characteristics for accurate and reliable complaint analysis.

D. Algorithms

1) *Multi-Task RoBERTa:* The Multi-Task RoBERTa architecture for complaint analysis integrates semantic and contextual understanding by combining the representation learning power of transformer-based models with task-specific classifiers. The shared RoBERTa encoder enhances feature consistency and improves classification reliability under various complaint formats where the transformer learning extracts discriminative textual features.
2) *RoBERTa:* RoBERTa is an effective model that learns the contextual patterns of complaint text by applying deep transformer layers to extract hierarchical textual features. With the incrementally contextual representation learning and organized architecture, RoBERTa improves both the classification stability and generalization ability and ensures reliable performance of complaint analysis.
3) *Task-Specific Classifiers:* Each task has a dedicated classifier head consisting of linear layers with dropout for regularization. The Aspect Classifier identifies the financial product category, the Severity Classifier determines the urgency level, the Emotion Classifier detects the customer's emotional state, and the Sentiment Classifier classifies the overall tone of the complaint.

E. Training and Optimization

The training process incorporates several techniques to ensure efficient and effective learning:

1) *Weighted Sampling:* Balances class distribution during training to handle class imbalance.
2) *Focal Loss:* Focuses on hard-to-classify samples with gamma set to 2.0, improving performance on minority classes.
3) *Gradient Clipping:* Prevents exploding gradients during backpropagation.
4) *Early Stopping:* Prevents overfitting with patience of 2 epochs.
5) *Linear Scheduler:* Adjusts learning rate for better convergence with warmup steps.
6) *Learnable Task Weights:* Automatically balances task contributions during training using learnable weights.

F. Integration of Explainability & Streamlit Framework

To increase the transparency and interpretability of complaint analysis, confidence scores and probability distributions were employed. Visual explanation methods highlight the most influential textual features for prediction decisions. The improved analytical transparency and trustworthiness during complaint analysis tasks is further strengthened by interpretability support that helps users to better understand the behavior of the prediction model. Visualization outputs also provide the advantage of discovering textual attributes that most affect the categorization result, which are relevant to complaint understanding.

The system was developed as a Streamlit-based web application. By connecting the user interface services with the complaint analysis module, a web application platform was built that provides features such as text input, file upload, real-time analysis, and prediction display. It provides an interactive environment for uploading complaint text or having real-time analysis and monitoring of complaint categories. The web application also enhances the user experience of developing smart complaint analysis applications, especially the ones requiring real-time monitoring and complaint understanding.

IV. EXPERIMENTAL RESULTS

1) *Accuracy:* The aggregate percentage of correctly categorized complaint samples among all the predictions made during evaluation is known as accuracy. It shows how well the complaint analysis framework works overall in accurately identifying aspects, severity, emotions, and sentiments across a variety of complaint categories.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

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 VII July 2026- Available at www.ijraset.com

2) *F1-Score:* The harmonic mean of precision and recall that provides a balanced measure of a model's performance by considering both false positives and false negatives is known as the F1 Score. It shows how well the complaint analysis framework maintains a balance between accurately identifying relevant complaints and avoiding false predictions, which is particularly important for imbalanced datasets where certain complaint categories may be underrepresented.

$$F1\,Score = 2 * \frac{Precision \text{ X } Recall}{Precision + Recall} \qquad (2)$$

Table.1 Performance Evaluation

Metric	Score
Overall Accuracy	91.88%
Overall F1 Score	88.28%

The overall performance of the Multi-Task RoBERTa model is summarized in Table 1. The system achieved an impressive overall accuracy of **91.88%** and an overall F1 score of **88.28%**, demonstrating the effectiveness of the multi-task learning approach for financial complaint analysis. These results validate that the shared RoBERTa architecture enables effective knowledge transfer across tasks, leading to robust and reliable predictions.

Task	Accuracy	F1 Score (Weighted)
Aspect	76%	65%
Severity	92%	90%
Emotion	100%	100%
Sentiment	96%	100%

The detailed performance metrics for each of the four tasks are presented in Table 2

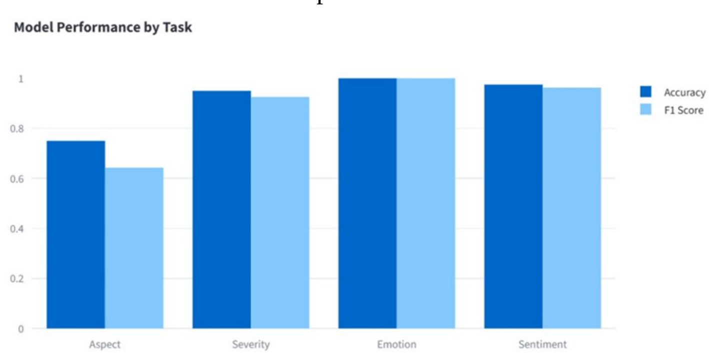


Fig.2 Model Performance by Task

The interactive Streamlit dashboard enables real-time analysis of complaint text. Fig. 3 shows the real-time analysis interface with navigation options.

Fig.3 Dashboard Navigation Interface

To demonstrate the system's practical utility, we tested it on a sample complaint: *"I was charged an overdraft fee of $35 even though I had sufficient balance."* The system produced the following predictions,

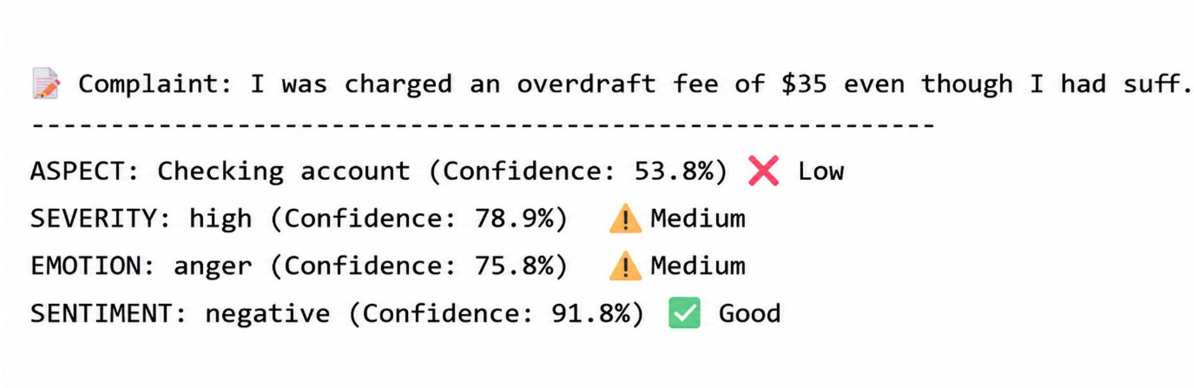


Fig.4 Real-Time Analysis Interface

The system correctly identified the complaint as related to a checking account with high severity. The customer's emotion was correctly detected as anger, and the sentiment was accurately classified as negative. The high confidence in sentiment prediction (91.8%) indicates that the model is particularly confident about negative sentiment classification. The aspect classification had the lowest confidence (53.8%), suggesting the model was less certain about the product category.

V. CONCLUSION

This project successfully developed and implemented a Multi-Task Financial Complaint Analysis System that leverages a shared RoBERTa architecture to simultaneously perform four classification tasks: aspect classification, severity prediction, emotion detection, and sentiment analysis. The system was trained and evaluated on the CFPB Consumer Complaint Database, achieving an overall accuracy of 91.88% and an F1 score of 88.28%, with per-task performance of 76% for aspect classification, 92% for severity prediction, 100% for emotion detection, and 96% for sentiment analysis. The interactive Streamlit dashboard enables real-time complaint analysis with confidence scores and probability distributions, making the system practical for real-world deployment in financial customer service environments. The results validate the effectiveness of the multi-task learning approach, demonstrating that a single shared RoBERTa model can effectively handle multiple related tasks while maintaining high performance across all of them. By addressing the limitations of existing single-task systems and providing a comprehensive, multi-dimensional analysis, this project contributes a valuable, reproducible, and offline-capable solution that can help financial institutions automate complaint processing, prioritize urgent issues, and gain data-driven insights for improving products and services, ultimately making financial customer service more responsive, efficient, and customer-centric through the thoughtful application of artificial intelligence.

VI. FUTURE SCOPE

The system can be extended to support multiple languages using multilingual models like XLM-RoBERTa, enabling global deployment across diverse financial institutions. Integration with existing Customer Relationship Management (CRM) systems and the development of a RESTful API would facilitate seamless enterprise deployment and real-time complaint processing. Additionally, incorporating explainable AI techniques such as LIME and SHAP would enhance transparency and trust by providing visual insights into model predictions and decision-making processes.

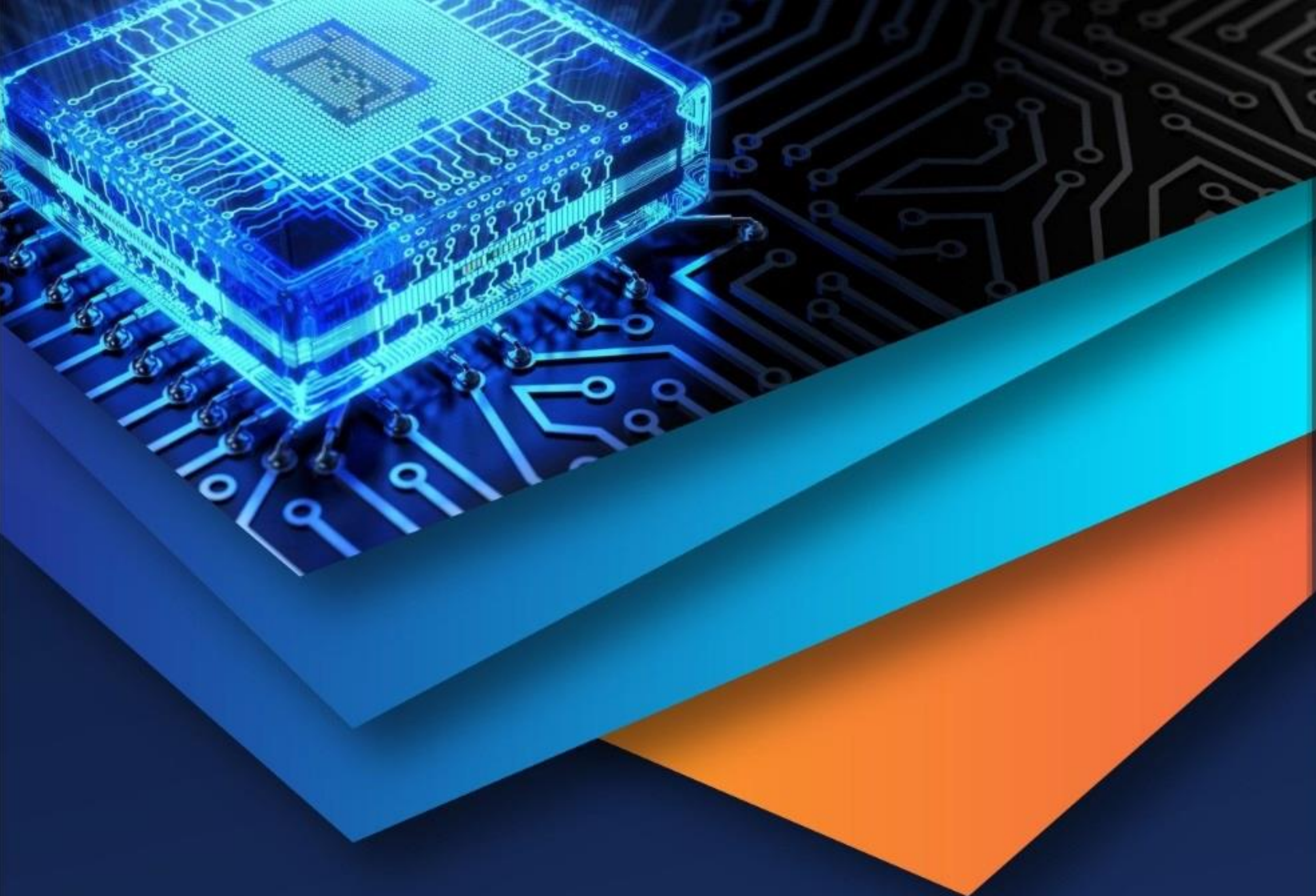
VII. ACKNOWLEDGEMENT

I express my sincere gratitude to my guide, Dr. B. Kranthi Kiran, Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University Hyderabad, for his valuable guidance, continuous support, and constructive suggestions throughout the development of this research work.

REFERENCES

[1] A. Vaswani et al., "Attention is All You Need," in Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008, 2017.
[2] A. Singh, R. Bhatia, and S. Saha, "Complaint and Severity Identification from Online Financial Content," IEEE Transactions on Computational Social Systems, vol. 11, no. 3, pp. 4130–4141, 2024.

[3] A. Joshi, A. Kumar, and N. Patel, "Leveraging Social Media and Multitask Natural Language Processing to Detect Financial Consumer Complaints," Economic Computation and Economic Cybernetics Studies and Research, vol. 59, no. 3, 2025.
[4] P. Jain et al., "An intelligent transformer based framework for bilingual financial complaint classification," Scientific Reports, vol. 16, 2026.
[5] S. Das et al., "When Words Can't Capture It All: Towards Video-Based User Complaint Text Generation," in ACM CIKM, 2025.
[6] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint arXiv:1907.11692, 2019.
[7] S. Minaee et al., "Large Language Models: A Survey," arXiv preprint arXiv:2402.06196, 2024.
[8] C. W. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," Computers & Education: Artificial Intelligence, vol. 2, p. 100033, 2021.
[9] S. Das, A. Singh, S. Saha, and A. Maurya, "Negative Review or Complaint? Exploring Interpretability in Financial Complaints," IEEE Transactions on Computational Social Systems, 2024.
[10] S. Das et al., "Deciphering the Complaint Aspects: Towards an Aspect-Based Complaint Identification Model with Video Complaint Dataset in Finance," in IEEE/CVF WACV, 2025, pp. 7195–7204.
[11] S. Das et al., "ExpertMix: Aspect and Severity Detection in Conversational Complaints," in Advances in Information Retrieval, 2026.
[12] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in Proceedings of IEEE ICCV, 2017, pp. 618–626.
[13] Z. Wang et al., "From Complaint Narratives to Monetary Relief: A Hybrid Machine Learning Framework for CFPB Consumer Complaints," arXiv preprint arXiv:2606.22664, 2026.
[14] S. Das et al., "Knowing What and How: A Multi-modal Aspect-Based Framework for Complaint Detection," Advances in Information Retrieval, 2023.
[15] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in *Proceedings of ACM SIGKDD*, 2016, pp. 1135–1144.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY