Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Sentiment Analysis and Smart Review of Stakeholder Comments in the eConsultation Module

B. Harish Goud
Department of Information Technology
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Shravya Jallepally
Department of Information Technology
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Ashish Pathak
Department of Information Technology
Chaitanya Bharathi Institute of Technology
Hyderabad, India

Shiva Karthik Bairy
Department of Information Technology
Chaitanya Bharathi Institute of Technology
Hyderabad, India

***Abstract*—The Ministry of Corporate Affairs employs the "eConsultation Module" to seek inputs from various stakeholders for their feedback and suggestions on proposed legislation or proposed amendments to that legislation. While the tool enables citizen participation, a vast number of comments can be collected during the consultation process in a matter of days. There is a high probability of overlooking any essential feedback, suggestions, opinions, etc., during this entire exercise. Thus, automation of the process becomes necessary, and we offer an AI-driven solution – Avalokan – that will analyze the stakeholder feedback efficiently. Our solution employs sentiment analysis to classify stakeholder feedback into either positive, negative, or neutral with respect to the proposed amendments, uses SentenceBERT cosine similarity technique to detect duplication of comments, and offers a version-based trend engine to track the evolution of sentiments with each passing version. The system gives an accuracy of 83.33% with F1-score of 0.7622 for real stakeholder comments. Additionally, the percentage of negative sentiments is consistently decreasing from 60% in v1.0 to 46% in v3.0, a measure of improvement of iterative policy-making. The performance of our proposed model beats a baseline of TF-IDF + Logistic Regression in accuracy.**

***Index Terms*—Sentiment Analysis, eConsultation, Government Policy, Natural Language Processing, Transformer Models, Decision Support System**

I. INTRODUCTION

Public participation lies at the heart of modern democratic governance. The Ministry of Corporate Affairs (MCA), India, has taken a significant step in this direction by deploying an eConsultation module that invites stakeholders, ranging from industry experts and legal professionals to ordinary citizens, to comment on proposed amendments and draft legislation [1], [2]. While such platforms strengthen the transparency and legitimacy of the legislative process, their very success creates a formidable operational challenge: thousands of responses can arrive within days of a consultation opening, quickly overwhelming the capacity of human reviewers to process them thoroughly.

The problem is not merely one of volume. Manual review is inherently slow, error-prone, and difficult to scale [3]. As the number of stakeholders grows, the depth of analysis that any fixed team of reviewers can provide remains essentially unchanged, creating a widening gap between the richness of public input and the quality of insight that actually informs policy decisions. Beyond scalability, human reviewers are susceptible to cognitive fatigue and analytical bias: a well-articulated but superficial comment may receive more weight than a technically insightful one that is poorly worded.

Addressing this gap calls for an intelligent, automated system that can process large volumes of unstructured text, classify the sentiment expressed in each submission, detect redundant or near-duplicate comments, and track how public opinion evolves across successive drafts of a policy. Several prior systems have applied BERT-based models to sentiment classification in legal and governance domains [10], [11]; however, they have largely treated sentiment analysis as a standalone task, without integrating duplicate detection, version-wise trend analysis, and a production-grade deployment pipeline into a single coherent platform. The system proposed in this paper *Avalokan* addresses this gap by combining DistilBERT-based sentiment classification, SentenceBERT cosine similarity for deduplication, a multi-version trend engine, and a full-stack Flask–MongoDB deployment into one unified decision support tool designed specifically for the eConsultation use case.

By applying Natural Language Processing techniques, the Ministry can transform thousands of paragraphs of unstructured feedback into structured, actionable intelligence making the consultation process not only more efficient but also more equitable, since every submission is processed with equal rigor regardless of how it is written.

The main contributions of this work are as follows:

- Developing an automatic sentiment analysis system for stakeholder comments on a large scale.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

- Implementing semantic similarity algorithms through SBERT for duplicate comment detection.
- Sentiment trend analysis of different versions of a policy draft.
- Full-stack implementation using Flask and MongoDB.

II. LITERATURE SURVEY

Large-scale textual feedback has been analyzed automatically, emerging as a fundamental area of research that developed alongside improvements to e-governance technology. Early sentiment analysis systems were based on conventional machine learning methods – significantly, combinations of TF-IDF feature extraction with Random Forest classifiers and Support Vector Machines – and were competitive in classification accuracy on domain-specific corpora [9], [10]. Such techniques are still computationally light and interpretable, making them convenient starting points for resource-constrained deployments.

Preprocessing is a precondition of any sentiment analysis pipeline, since real-world consultation text is disorderly and disjointed. Standard steps include tokenization, stop-word removal, lemmatization, and elimination of punctuations and special characters. Visualization tools such as word clouds and sentiment distribution charts assist policymakers in seeing key themes at a glance, without requiring strong technical know-how.

With the introduction of transformer-based language models, the landscape changed significantly. BERT [11] showed that deep bidirectional pretraining on large corpora generates contextual representations which significantly outperform previous practices on classification problems. Subsequent work expanded it to domain-specific variants: LegalBERT adapted the BERT architecture to legal text, and RoBERTa enhanced the training regime of BERT to generate more robust representations. The predicted sentiment probability under these models is computed as:

$$P(y \mid x) = \text{softmax}(W \cdot h_{[\text{CLS}]} + b) \tag{1}$$

Later developments moved towards hybrid architectures that combine the representational power of transformer embeddings with the decision-boundary versatility of classical classifiers. [10] suggested combining BERT embeddings with a Random Forest classifier and found enhanced generalization on skewed sentiment datasets – a result directly applicable to the current study, due to class imbalance in the eConsultation corpus. Similarly, Xie et al. [12] used BERT in combination with FastText to analyze sentiment in educational text, demonstrating that lightweight auxiliary encoders can offset domain vocabulary gaps that a general-purpose BERT model may not fully capture.

Sarcasm and implicit sentiment remain open challenges across all these architectures. Arif and Nayak [13] showed that even fine-tuned BERT models struggle with figurative language and indirect expressions of disapproval – a limitation particularly relevant in policy consultation settings, where stakeholders may signal resistance through rhetorical questions or qualified praise rather than explicit negative statements.

For duplicate and near-duplicate detection, Sentence-BERT [8] introduced a siamese network architecture producing semantically meaningful sentence embeddings that support efficient cosine similarity comparison. This method is far more scalable than cross-encoding every pair of submissions, making it suitable for large-volume consultation scenarios.

For keyword extraction, TF-IDF frequency analysis has been widely used to surface recurring themes across large document sets, allowing policymakers to identify the most frequently raised concerns. Hybrid models combining RoBERTa or LegalBERT with classical classifiers have shown strong performance in legal sentiment analysis, though at greater computational expense [9], [10].

On the intersection of sentiment analysis and civic technology, Simonofski et al. [5] studied how social media sentiment can complement conventional policymaking processes, identifying the need for tools that bridge the gap between unstructured public opinion and structured legislative workflows. This directly motivates the design of Avalokan as an end-to-end decision support system rather than a standalone classifier.

A keyword mapping algorithm has also been proposed to identify references to specific provisions within draft legislation [3], enabling finer-grained sentiment attribution. The Trend Engine concept – tracking sentiment shifts across drafts (v1, v2, v3) – provides a longitudinal view of how public reception evolves as policymakers incorporate stakeholder feedback, which is a core component of the system this paper describes.

III. System Architecture and Implementation

The proposed system is built upon a layered architecture consisting of three principal components: a React-based Frontend Interface, a Flask Backend, and an AI Processing Module. This separation of concerns allows each layer to be developed, tested, and scaled independently, which is essential when processing the large and unpredictable volumes of feedback characteristic of active public consultations.

Stakeholders interact with the frontend to submit comments on a draft policy. These submissions are relayed to the backend via RESTful API calls and persisted in a MongoDB database organized into four collections: Users, Policies, Drafts, and Analysis Results. MongoDB's document-oriented schema accommodates the variable structure of free-text comments without requiring rigid table definitions, and its horizontal scaling capability supports future growth in submission volume.

Once stored, each comment enters the AI Processing Module through a multi-stage preprocessing pipeline: (i) Unicode normalization and lowercasing to ensure encoding consistency; (ii) removal of URLs, email addresses, and special characters that carry no semantic value; (iii) contraction expansion (e.g., *won't* $\rightarrow$ *will not*) to reduce vocabulary fragmentation; (iv) tokenization using the DistilBERT WordPiece tokenizer, which handles out-of-vocabulary terms through subword decomposition; and (v) truncation or padding to a maximum

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

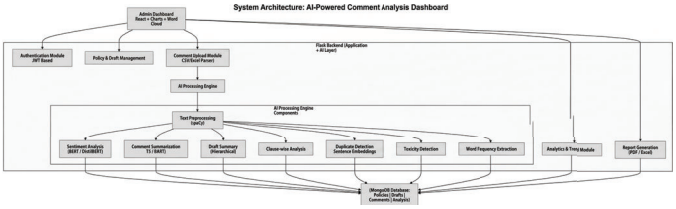

Fig. 1. Detailed AI Processing Flow and Modular Architecture.

sequence length of 512 tokens as required by the model. Stopword removal was deliberately omitted at the transformer input stage, since DistilBERT's attention mechanism is capable of learning to down-weight uninformative tokens internally.

The preprocessed token sequences are passed to the sentiment classification component. The system uses a pretrained DistilBERT model fine-tuned on SST-2 (`distilbert-base-uncased-finetuned-sst-2-english`), which produces a contextual embedding for the entire input via the special `[CLS]` token. The pretrained SST-2 model was adapted to a three-class setting through post-processing of prediction probabilities. A linear classification layer maps this embedding to sentiment logits, and a softmax activation converts them to a probability distribution over the three sentiment classes:

$$P(y \mid x) = \text{softmax}(W \cdot h_{[\text{CLS}]} + b) \tag{2}$$

where $h_{[\text{CLS}]}$ is the contextual embedding of the input sequence, and W and b are the learned parameters of the classification layer.

To detect duplicate or near-duplicate submissions a common occurrence in coordinated lobbying campaigns Sentence-BERT embeddings are computed for each comment. The semantic similarity between any two comments A and B is measured using cosine similarity:

$$\text{CosSim}(A, B) = \frac{A \cdot B}{\|A\| \, \|B\|} \tag{3}$$

Comment pairs whose cosine similarity exceeds a predefined threshold are flagged as duplicates and excluded from the primary sentiment aggregation, preventing organized submission campaigns from distorting the overall sentiment signal. A trend analysis module then aggregates the per-comment sentiment scores across all submissions for a given draft version, producing version-level sentiment distributions that can be compared across v1.0, v2.0, and v3.0 of a policy document.

A. Dataset Description

The dataset used in this study combines simulated and real-world stakeholder comments to balance controlled experimental conditions with real-world validity. The simulated portion was generated using a large language model prompted to produce realistic policy feedback across a range of sentiments, writing styles, and levels of technical detail. These synthetic comments were used exclusively during system development and pipeline validation, they were not included in the final performance evaluation reported in Section IV.

The real-world portion was collected through the Avalokan platform itself: registered users were invited to submit comments on three successive versions of a sample draft policy, mirroring the structure of an actual MCA consultation. All submissions were collected with user consent. Each comment was independently labelled by two annotators using a three-class schema: *Positive* (the commenter broadly supports the amendment or expresses approval), *Negative* (the commenter opposes or raises concerns), and *Neutral* (factual observations, questions, or ambiguous statements). Disagreements between annotators were resolved through discussion, with a third team member serving as a tiebreaker where consensus could not be reached.

The final dataset comprises approximately 1,400 comments distributed across the three classes: 32% Positive (≈448 comments), 47% Negative (≈658 comments), and 21% Neutral (≈294 comments). The pronounced negative skew reflects the typical pattern of policy consultations, where stakeholders who feel strongly enough to respond are more likely to raise objections than to express general approval. This class imbalance was accounted for in evaluation by reporting weighted precision and F1-score in addition to overall accuracy.

IV. RESULTS AND DISCUSSION

To establish a meaningful point of comparison, a classical baseline was constructed using TF-IDF weighted bag-of-words features combined with a Logistic Regression classifier trained with L2 regularization. This approach represents the standard pre-transformer pipeline for text classification and has been widely used as a reference model in sentiment analysis literature [9], [10]. The baseline was trained and evaluated on the same 150 real-world comments used for the proposed model, under identical train/test split conditions, to ensure a fair comparison.

The proposed system uses a DistilBERT model (`distilbert-base-uncased-finetuned-sst-2-english`) for sentiment classification of stakeholder comments. Experiments were conducted on approximately 150 real-world comments collected through the Avalokan platform, held out entirely from the training process and used solely for evaluation.

TABLE I
PERFORMANCE EVALUATION OF THE PROPOSED SYSTEM.

Model	Accuracy	F1 Score
TF-IDF + Logistic Regression (Baseline)	75%	0.74
Proposed DistilBERT Model	83.33%	0.762

The DistilBERT model achieved an overall accuracy of 83.33%, a weighted precision of 70.9%, recall of 83.33%, and an F1-score of 0.7622, compared to the baseline's accuracy of 75% and F1-score of 0.74. While the DistilBERT model surpasses the baseline on accuracy, the almost similar F1-score (0.7622 vs. 0.74) reflects the impact of class imbalance on per-class precision specifically, the model tends to over-predict the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

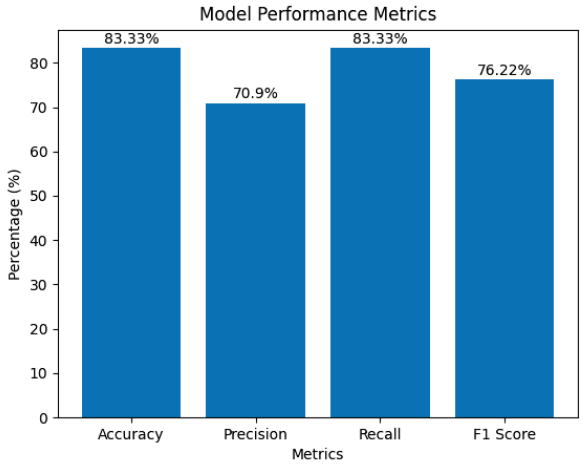

Fig. 2. Evaluation metrics of the sentiment classification model showing accuracy, precision, recall, and F1-score computed on real stakeholder comments.

majority Negative class, suppressing precision for the minority Positive and Neutral classes. This is a known behaviour of models fine-tuned on balanced benchmark datasets such as SST-2 when applied to skewed real-world distributions, and it motivates future fine-tuning directly on the eConsultation corpus.

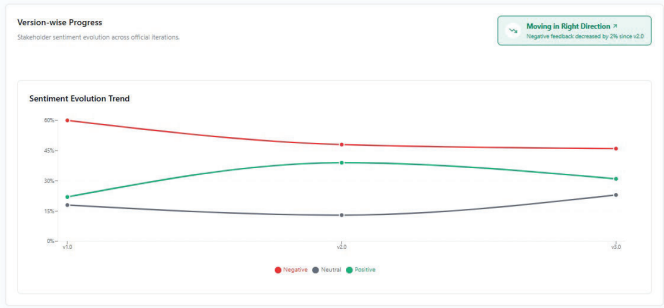

Fig. 3. Version-wise Sentiment Evolution Trend: Stakeholder sentiment distribution (Positive, Negative, Neutral) tracked across three official draft iterations (v1.0–v3.0). A consistent decrease in negative feedback and a general increase in positive sentiment validate the iterative policy refinement process.

The version-wise analysis provides a more encouraging picture of the system's practical utility. Applied to stakeholder comments across three successive draft versions, the system tracks a clear directional trend: negative sentiment decreased from 60% in v1.0 to approximately 46% in v3.0, while positive sentiment rose from 23% in v1.0 to a peak of 35% in v2.0 before stabilizing at around 30% in v3.0. Neutral sentiment remained relatively stable, ranging from 15% in v1.0 to approximately 23% in v3.0.

The slight dip in positive sentiment from v2.0 to v3.0, alongside the rise in neutral sentiment, suggests that while the v3.0 revisions successfully addressed a significant share of stakeholder concerns, some newly introduced provisions generated uncertainty or mixed reactions rather than outright approval. This kind of nuanced, version-level insight distinguishing between "concerns resolved" and "new ambiguities introduced" is precisely the type of intelligence that is difficult to extract through manual review but is surfaced naturally by the trend analysis module.

Overall, the system demonstrates that the sentiment trend is moving in a positive direction across draft iterations, with the consistent reduction in negative feedback serving as a quantitative indicator of iterative policy improvement.

V. LIMITATIONS

The greatest linguistic weakness of the existing system is the way it treats implicit and figurative sentiment. Stakeholder disagreement in policy consultations is often expressed through rhetorical questions, hedged phrasing, or sarcastic framing e.g., a remark like *I am sure this amendment will do the drafters of it a great service* carries negative sentiment that a superficial classifier is prone to misread as positive. The DistilBERT model in this case was fine-tuned on SST-2, a movie review sentiment dataset in which sentiment is predominantly explicit. Being accustomed to the more subtle register of legal and policy discourse, the model has no mechanism for identifying irony or situational sarcasm. Addressing this would require either a dedicated sarcasm pre-filtering step such as the fine-tuned BERT methodology investigated by Arif and Nayak [13] or domain-adaptive fine-tuning on annotated eConsultation data containing sarcastic examples.

The second constraint is related to scaling under high-load conditions. The present evaluation was carried out on approximately 150 real-world comments, which suffices for proof-of-concept validation but is orders of magnitude fewer than the volume a national-level MCA consultation would generate. The DistilBERT inference pipeline, though lighter than full BERT, still processes comments one by one in the current implementation. At scale e.g., 50,000 submissions arriving within 48 hours - this would generate intolerable latency without asynchronous batched inference, processing queues, or model quantization to minimize per-sample inference time. The MongoDB database is horizontally scalable, yet the AI module remains a bottleneck that must be addressed prior to production deployment at national scale.

Third, the system operates in near-real-time rather than actual real-time. The sentiment score is recomputed periodically in response to batch processing of comments, not instantly upon submission to the dashboard. This approach suffices under most consultation circumstances, but in emergencies requiring rapid regulatory response, or during fast-moving societal events, the processing delay can result in significant shifts in public opinion going unnoticed for minutes or even hours. Integrating the inference pipeline with a tool such as Apache Kafka or Redis could help reduce processing latency.

Moreover, although the evaluation dataset is of real-world origin, its collection was carried out through an artificial platform with a relatively homogeneous user community. Feedback gathered from a broader public survey involving

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

multiple regional languages, varying literacy levels, and diverse linguistic backgrounds may exhibit different distributional characteristics compared to the current corpus. The system's performance on such heterogeneous input has yet to be established.

VI. Societal Impact and Practical Implications

In its simplest form, Avalokan addresses a democratic deficit. This deficit is produced when a government consultation receives thousands of responses but only a fraction can be reviewed the voices of the majority of participants go unheard, no matter how committed the system claims to be to engagement. By automating sentiment classification, deduplication, and trend tracking, the system ensures that every submission is studied with the same rigor, regardless of how eloquently it is written or when during the consultation window it was submitted. This fairness of processing is not just a technological feature; it is a material improvement to the fairness of the legislative process.

To policymakers, the practical advantages are immediate and concrete. Instead of ministry officials being given a stack of thousands of raw submissions at the end of a consultation, they are presented with a dashboard displaying the overall sentiment distribution, the prevailing concerns as indicated by keyword analysis, the percentage of duplicate entries removed, and critically a version-wise trend chart showing how sentiment has shifted across past revisions. This compresses what would otherwise be weeks of manual post-consultation review into an hours-long analytical task, liberating officials to focus on policy content rather than the logistics of document processing.

The societal implications extend beyond efficiency. Automated deduplication deters coordinated lobbying in which one interest group submits hundreds of near-identical responses from artificially inflating the apparent weight of a particular viewpoint. Toxicity filtering removes abusive or offensive posts prior to review, safeguarding the integrity of the consultation record. Together, these features make the consultation process more resistant to manipulation and more reflective of genuine public sentiment, which strengthens the democratic legitimacy of the resulting legislation.

Beyond the MCA, the Avalokan architecture is domain-agnostic. Any organization that receives large volumes of structured public opinion environmental regulators, municipal planning authorities, public health agencies faces the same information overload problem. The modular design of the system, with clearly separated preprocessing, classification, deduplication, and trend analysis components, means it can be adapted to new domains by retraining the classification layer on domain-relevant data while leaving the rest of the pipeline intact. This generalizability significantly extends the potential societal contribution of the work.

VII. CONCLUSION

In this paper, Avalokan an AI-based decision support system meant to transform the manner in which government ministries process and respond to public input in legislative consultations was introduced. The system combines DistilBERT-based sentiment classification, Sentence-BERT cosine similarity for duplicate detection, multi-version trend analysis, and a full-stack Flask–MongoDB implementation into a single unified platform an integration which, to the best of our knowledge, has not been previously demonstrated in the eConsultation domain.

Evaluated on 150 real-world stakeholder comments, the system achieved an accuracy of 83.33% and an F1-score of 0.7622, outperforming the TF-IDF + Logistic Regression baseline on precision while highlighting the class imbalance issue a well-known challenge when benchmarked models are applied to skewed real-world distributions. The version-wise sentiment trend findings confirmed the system's capacity to identify meaningful shifts in public opinion across drafts, with positive sentiment declining from 60% in v1.0 to 46% in v3.0 a quantitative signal of iterative policy improvement that would be practically impossible to extract through manual inspection.

The identified limitations sarcasm detection, sequential inference bottlenecks, near-real-time rather than streaming updates, and dataset diversity establish a clear and actionable research agenda. Future work will focus on domain-adaptive fine-tuning of the classification model on a larger and more diverse eConsultation corpus; a streaming inference pipeline to enable truly real-time sentiment updates; and adoption of a blockchain-based immutable audit ledger to ensure the integrity and transparency of the entire consultation record. These directions build directly on the foundation the current system has established, collectively advancing Avalokan from a proof-of-concept towards a production-grade tool for democratic governance at scale.

ACKNOWLEDGMENT

The authors wish to thank the Department of Information Technology at Chaitanya Bharathi Institute of Technology, Hyderabad, for providing the infrastructure and academic environment that made this work possible. We are also grateful to all the participants who submitted comments through the Avalokan platform during data collection, and to the annotators whose careful labelling effort formed the foundation of the evaluation. Finally, we thank our peers and reviewers whose feedback helped sharpen the ideas presented in this paper.

REFERENCES

[1] A. Macintosh, "Characterizing e-participation in policy-making," in *Proc. 37th Hawaii Int. Conf. System Sciences*, IEEE, 2004.
[2] R. Deshmukh et al., *E-governance and Digital Innovation*. Springer, 2024.
[3] Z. Jin and R. Mihalcea, *Natural Language Processing for Policy-Making*. Springer, 2023.
[4] M. Mansoor et al., "Semantic similarity detection," 2020.
[5] A. Simonofski et al., "Policy-making with social media," 2021.
[6] E. Cambria et al., *A Practical Guide to Sentiment Analysis*. Springer, 2017.
[7] S. Poria et al., "Multimodal sentiment analysis," 2017.
[8] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using siamese BERT-networks," *arXiv preprint arXiv:1908.10084*, 2019.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

[9] E. Demir and M. Bilgin, "Sentiment analysis from Turkish news texts with BERT-based language models and machine learning algorithms," 2023.
[10] U. Ghosh, S. Sarkar, S. Jana, I. Bhattacharya, K. Singh, and P. Kumari, "A hybrid framework for sentiment analysis on textual data using BERT embeddings and random forest classifier," 2026.
[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[12] P. Xie, H. Gu, and D. Zhou, "Modeling sentiment analysis for educational texts by combining BERT and FastText," 2024.
[13] M. F. Arif and J. Nayak, "Fine-tuned BERT model for accurate hate speech detection in social media," 2024.