ISSN No. 0976-5697

VOLUME 17 SPECIAL ISSUE 1, JULY 2026

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

LegisAI: AI-Powered Legislative Comment Analysis

Dr. Ruchi Vyas
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan, India
ruchi.vyas@gits.ac.in

Dr. Mayank Patel
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan, India
mayank.patel@gits.ac.in

Sanvi Purohit
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan, India
snvpurohit@gmail.com

Riya Mittal
Geetanjali Institute of Technical Studies
Udaipur, Rajasthan, India
riya.mittal0913@gmail.com

Abstract: The increasing volume of stakeholder feedback in legislative processes makes manual analysis inefficient and error-prone. This paper presents LegisAl, an Al-based system designed to automatically analyze legislative comments using Natural Language Processing techniques. The system performs sentiment analysis to classify opinions, generates concise summaries of lengthy comments, and visualizes frequently used keywords through word clouds. Transformer-based models such as DistilBERT and DistilBART are utilized to achieve efficient and accurate processing. The proposed solution significantly reduces manual effort and enhances decision-making by providing structured insights. Experimental results demonstrate an approximate accuracy of 85% in sentiment classification with effective summarization performance.

Keywords: Natural Language Processing, Sentiment Analysis, Text Summarization, Word Cloud, Transformer Models, Legal AI

I. INTRODUCTION

In public consultation processes, draft legislations are often released for a specific period during which stakeholders can submit their comments and suggestions. With the increasing use of digital platforms such as eConsultation modules, a large volume of feedback is received from various stakeholders. Manually reviewing and analyzing such a high number of comments becomes time-consuming and may lead to certain suggestions being overlooked or insufficiently analyzed.

To address this challenge, automated text analysis techniques can be applied to efficiently process stakeholder feedback. One such technique is sentiment analysis, a Natural Language Processing (NLP) method used to identify and classify opinions expressed in text as positive, negative, or neutral. By applying sentiment analysis to stakeholder suggestions, it becomes easier to understand the overall opinion trends and individual sentiments regarding draft legislation.

In this work, we propose an AI-based solution that utilizes Natural Language Processing and Machine Learning techniques to analyze stakeholder feedback submitted in the eConsultation system. The proposed system performs sentiment classification, summary generation, and word cloud visualization to highlight frequently used keywords in the suggestions. These features help in quickly understanding the key concerns, opinions, and patterns in the feedback received.The system leverages key technologies including DistilBERT for sentiment classification, DistilBART for text summarization, and word cloud generation for visualization. These technologies enable efficient processing and interpretation enable efficient.

II. TECHNOLOGIES USED

LegisAI is designed by integrating multiple advanced Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies to ensure efficient, accurate, and scalable text processing of legislative comments. At the core of the system is DistilBERT, a lightweight and optimized transformer model derived from BERT. It is specifically utilized for sentiment analysis, where it classifies stakeholder comments into sentiment categories such as positive, negative, or neutral. Despite being computationally efficient, DistilBERT maintains high accuracy and enables faster inference, making it highly suitable for real-time or large-scale applications.

For text summarization, the system employs DistilBART, a compressed variant of the BART model that is capable of performing abstractive summarization. Unlike extractive methods, DistilBART generates concise and meaningful summaries by understanding the context of the input text and recreating it in a shorter form while preserving its original intent. This feature is particularly useful when dealing with lengthy legislative documents or large volumes of stakeholder feedback.

In addition to these models, various techniques from Natural Language Processing are incorporated to enhance data quality and processing efficiency. These techniques include tokenization, text cleaning, stop-word removal, and normalization, which collectively transform raw textual data into a structured format suitable for analysis. NLP plays a crucial role in ensuring that the input data is consistent and free from noise, thereby improving the performance of downstream models.

To further support interpretability and insight extraction, the system integrates word cloud visualization, a graphical technique that highlights the most frequently occurring words within the dataset. This allows users to quickly identify dominant themes, patterns, and key concerns expressed in stakeholder comments without reading each entry individually. The entire system is implemented using Python, which serves as the primary programming language for development and integration. Python provides extensive libraries and frameworks for machine learning, NLP, and data visualization, enabling seamless coordination between different components of the system. Its flexibility and strong community support make it an ideal choice for building a robust and scalable solution like LegisAI.

III. LITERATURE SURVEY

Several studies have explored sentiment analysis using machine learning and deep learning approaches for analyzing textual data. Gupta and Kumar proposed sentiment classification using multiple models to monitor customer reviews, demonstrating improved accuracy with ensemble techniques.[1] Kumar et al. provided a comprehensive review of evolving sentiment analysis techniques including traditional machine learning, deep learning, and transformer-based models. Similarly, Suryawanshi analyzed various sentiment analysis approaches such as SVM, Naïve Bayes, CNN, and LSTM, highlighting their applications in opinion mining and social media analysis. Although these approaches improved sentiment classification accuracy, they primarily focused on sentiment detection and did not integrate summarization or visualization features.[2][3]

Several works focused on sentiment analysis applications and computational techniques in NLP. Jena introduced fundamental sentiment analysis workflow including preprocessing, feature extraction, and classification techniques.[4] Raj et al. performed sentiment analysis on news datasets to analyze public opinion trends using AI-based models. IEEE research also presented various computational methods including lexicon-based and deep learning techniques for sentiment classification. These studies improved understanding of sentiment extraction from textual data; however, they lacked integration with automated summarization and keyword visualization.[5]

Recent advancements in text summarization using deep learning models have shown promising results. Krishna et al. proposed a deep learning-based text summarization system using encoder-decoder architecture for automated news digest generation.[6] Another study evaluated large language models using prompt engineering techniques for generating concise summaries. These models effectively generated meaningful summaries but were limited to summarization tasks and did not include sentiment analysis.[7]

Additional research focused on computational approaches and AI-based text analysis tools. Benni demonstrated Python-based computational analysis for processing datasets efficiently.[8] Wizrai discussed AI text analysis tools for extracting insights such as sentiment detection, keyword extraction, and summarization. However, these studies discussed general analysis techniques and did not propose an integrated system combining sentiment analysis, summarization, and visualization.[9]

These studies highlight that while individual components such as sentiment analysis, summarization, and text analytics exist, there is a lack of an integrated framework combining these features. The proposed LegisAI system aims to address this limitation by developing a unified platform for sentiment analysis, summarization, and visualization.[10]

H. Gupta and M. Patel presents a comprehensive experimental study that compares extractive text summarization techniques with topic modeling approaches, both of which are important tasks within Natural Language Processing (NLP). Topic modeling focuses on identifying and extracting meaningful themes or topics from large textual datasets, thereby enabling better understanding and organization of unstructured text. In this work, the authors employ Latent Semantic Analysis (LSA), a widely used statistical method for uncovering hidden relationships between words and documents. LSA is implemented using truncated Singular Value Decomposition (SVD), which helps reduce the dimensionality of the text data while preserving its most significant semantic structures.

IV. IMPLEMENTATION

The proposed system follows a structured pipeline for processing legislative comments.

Algorithm:

1. Input legislative comments (single or dataset)

2. Perform text preprocessing (cleaning and normalization)

3. Apply sentiment analysis using DistilBERT
4. Generate summary using DistilBART

5. Create word cloud visualization

6. Display final results

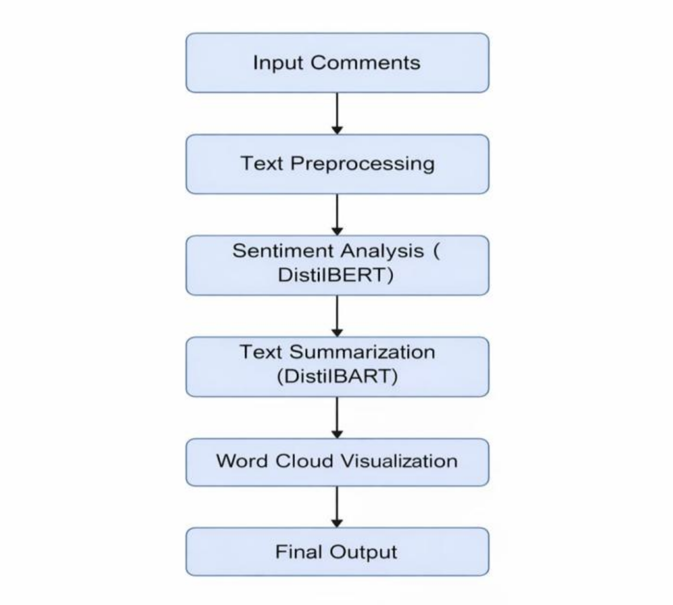

Figure 1: System Flowchart of LegisAI

The proposed system follows a structured and systematic pipeline designed for efficient processing and analysis of legislative comments. Initially, the system accepts legislative comments as input, which may consist of either a single comment or a large dataset of comments collected from various sources. Once the input is provided, the system performs text preprocessing, which includes essential steps such as data cleaning, removal of noise (like special characters and stop words), and normalization to ensure consistency in the textual data. This preprocessing stage enhances the quality of the input data and prepares it for further analysis.

Subsequently, the processed text is subjected to sentiment analysis using the advanced transformer-based model DistilBERT, which efficiently classifies the sentiments expressed in the comments into categories such as positive, negative, or neutral. Following sentiment detection, the system employs another powerful model, DistilBART, to generate concise and meaningful summaries of the legislative comments, thereby enabling quick understanding of large volumes of textual information.

In addition to textual analysis, the system also incorporates a visualization component by generating a word cloud, which highlights the most frequently occurring terms in the dataset and provides an intuitive graphical representation of key themes. Finally, all the processed outputs, including sentiment results, summarized content, and visualizations, are displayed to the user in an organized manner, facilitating easy interpretation and informed decision-making.

V. Results And Analysis

The system was tested on multiple legislative comments to evaluate its performance.

The sentiment analysis model achieved an approximate accuracy of 85% indicating reliable classification performance.

The summarization model generated concise summaries that preserved the core meaning of the original comments. The word cloud visualization effectively highlighted frequently used keywords, allowing quick identification of major topics discussed by stakeholders.

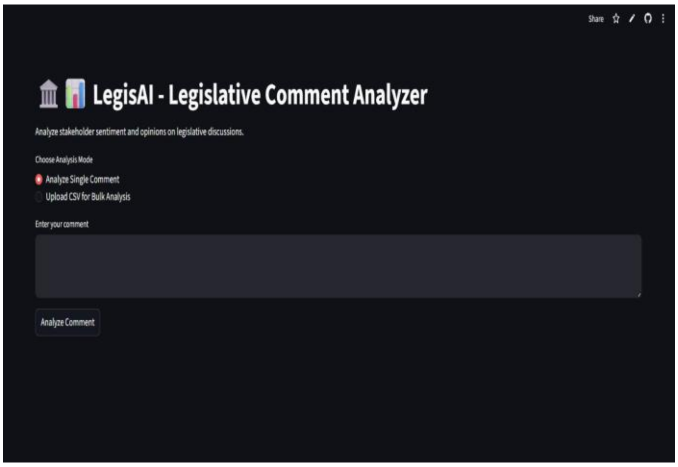

Figure 2: User Input for Comment Input or Bulk Comment Analysis

Figure 3: Word Cloud Visualization Output

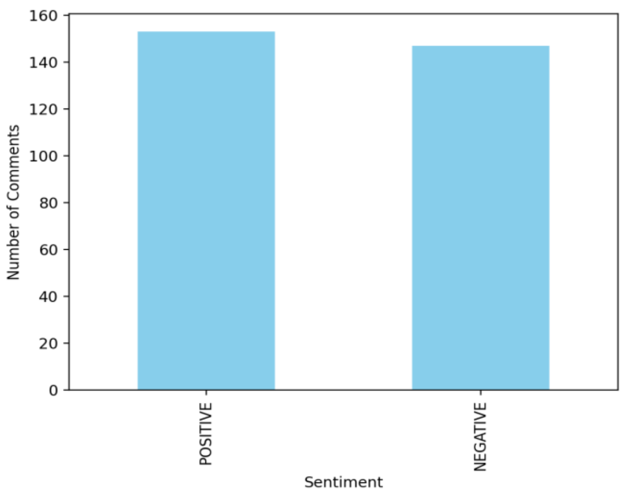

Figure 4: Sentiment Distribution

VI. CONCLUSION AND FUTURE WORK

This paper presented LegisAl, an Al-based system for analyzing legislative comments using sentiment analysis, summarization, and word cloud visualization. The system significantly reduces manual effort and provides structured insights from large volumes of data.

The results demonstrate that transformer-based models can effectively handle text analysis tasks with good accuracy and efficiency. The integration of multiple features into a single system makes LegisAl practical for real-world applications.

In the future, the system can be enhanced by integrating advanced large language models for better contextual understanding, supporting multilingual input, and enabling real-time analysis. Additional features such as topic modeling and clustering can further improve analytical capabilities.

VII. REFERENCES

[1] Gupta, C. P., & Kumar, V. V. R(2025). Sentiment Analysis : Using different models for monitoring and analyzing customer reviews. Indian Journal of Marketing, 55(5),8. https://doi.org/10.17010/ijom/2025/v55/i5/175017

[2] Kumar, M., Khan, L., & Chang, H. (2025). Evolving techniques in sentiment analysis: a comprehensive review. PeerJ Computer Science, 11, e2592. https://doi.org/10.7717/peerj-cs.2592

[3] Suryawanshi, N. S. (2024). Sentiment analysis with machine learning and deep learning: A survey of techniques and applications. International Journal of Science and Research Archive, 12(2), 005–015. https://doi.org/10.30574/ijsra.2024.12.2.1205

[4] Jena, A. (2025a, August 18). Introduction to Sentiment Analysis in natural language processing (NLP). Knowledge Tank. https://www.projectguru.in/introduction-to-sentiment-analysis-in-natural-language-processing-nlp/

[5] Benni, S. D. (2024). Green Fuel Synthesis from Jatropha curcas Seed Oil: A Computational Approach using Python Programming Language. Journal of Electrical Systems, 20(11s), 3388–3396. https://doi.org/10.52783/jes.8092

[6] Krishna, K. M. R., Somasundaram, K., Arulmozhivarman, P., Immanuel, S. A., & Rajkumar, E. R. (2025). Deep learning for text summarization using NLP for automated news digest. Scientific Reports, 15(1), 36343. https://doi.org/10.1038/s41598-025-20224-1

[7] An evaluation of large language models on text summarization tasks using prompt engineering techniques. (n.d.). https://arxiv.org/html/2507.05123v1

[8] Raj, R. K. M., Vochozka, M., & Sujatha, S. (2025). Sentiment analysis of news: unveiling AI's role in sustainability and no poverty (SDG1). Discover Sustainability, 6(1). https://doi.org/10.1007/s43621-025-01456-7

[9] Sentiment Analysis: a comprehensive study of computational techniques, applications, and datasets. (2024, December 16). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/10860172

[10] Wizrai. (2026, January 30). AI Text Analysis: 6 Best Tools, Techniques & Use Cases [2025]. Wizr AI | Build Autonomous Enterprises Powered by AI. https://wizr.ai/blog/ai-text-analysis-tools-techniques-use-cases/