IES International Journal of Multidisciplinary Engineering Research

Sentiment Analysis on Social Media Using Deep Learning Techniques

Alen Shaju[1], Hamis Ahammed K[2], Jismi Joy[3], Vysagh A U[4], Aysha[5]

[1,2,3,4]Student, CSE, IES College of Engineering, Kerala

[5]Assistant Professor, CSE, IES College of Engineering, Kerala

Email_id : jismijoy12@gmail.com, hamisahmd818@gmail.com, vysaghau@gmail.com, alenshaju524@gmail.com, aysha@iesce.info

Abstract

Social media platforms generate vast amounts of opinion-rich data that reflect public emotions and attitudes. The project explores sentiment analysis using deep learning models to classify emotions in user-generated text. The system achieves high accuracy in identifying emotional polarity, helping interpret public opinion effectively. The project uses YouTube Data API, social media posts to extract comments from video links posted on social media. Comments are cleaned and passed through a multilingual sentiment analysis model based on BERT to classify emotions. The model supports both English and regional language inputs, including code-mixed text. Visual results are displayed as pie charts along with a comment-wise sentiment summary. The tool enables real-time sentiment extraction from video discussions using deep learning.

Keywords: Sentiment Analysis, Social Media, Deep Learning, Emotion Classification, YouTube Data API, Real-Time Sentiment Extraction, Public Opinion.

DOI: https://doi.org/10.5281/zenodo.18495620

1. Introduction

In today’s hyper-connected digital landscape, public opinion and sentiment, particularly within dynamic social media ecosystems like Twitter and YouTube, exert a powerful influence on corporate strategy, political discourse, and market trends. Organizations increasingly rely on real-time feedback from these platforms for crisis management, product development, and understanding consumer perception. However, the sheer volume, velocity, and linguistic complexity of social media text—characterized by slang, acronyms, and sarcasm—overwhelm traditional sentiment analysis methods, leading to delayed or inaccurate insights. This limitation can severely hamper an organization's ability to react promptly and appropriately to critical shifts in public mood.

The BERT-Based Cross-Platform Sentiment Analysis System for Twitter and YouTube project addresses this challenge by leveraging advanced deep learning to create an intelligent, real-time opinion mining framework. Our system allows a user to select a social media platform and input a specific post or video URL, upon which it automatically extracts and analyzes the associated textual content (tweets or comments).

By employing the BERT (Bidirectional Encoder Representations from Transformers) model, which is superior at contextual language understanding, the system goes beyond simple keyword matching to accurately classify the latent sentiment (Positive, Negative, Neutral) in complex social media text. The system then aggregates these individual scores to provide a comprehensive, actionable summary, visually presented via a pie chart on a user interface.

1.1. Sentiment Analysis (BERT Model)

The core analytical component, the Sentiment-Engine, utilizes the fine-tuned BERT (Bidirectional Encoder Representations from Transformers) deep learning model to perform robust, contextual analysis of the preprocessed social media text. It accurately classifies the sentiment of individual pieces of content (tweets or comments) into discrete categories (Positive, Negative, Neutral), effectively handling the nuances, slang, and sarcasm prevalent in online language. This engine provides the objective sentiment score necessary for the final visualization.

1.2. Data Acquisition Module

This module is responsible for secure and efficient interaction with external platforms. It handles API calls and scraping logic to reliably retrieve the high volume of text data associated with the user-specified Twitter or YouTube URL. It manages API authentication, rate limits, and extracts the raw textual content (e.g., tweet text, YouTube comments, and related metadata), ensuring a steady and clean stream of information for downstream processing.

1.3. Visualization and Summarization Module

This integrated component transforms raw analytical results into actionable intelligence. It aggregates the individual sentiment classifications into a unified score and displays the final distribution of Positive, Negative, and Neutral sentiment via a clear, intuitive pie chart. This module also leverages the BART summarization capability to provide a concise textual overview of the content, enhancing user comprehension and decision-making speed.

1.4. Multimodal and Localization Modules

This layer adds advanced features and practical utility to the system. The CLIP (Contrastive Language-Image Pre-training) component handles the visual analysis of any associated images or video thumbnails to incorporate multimodal context into the analysis. Additionally, the system includes an optional capability to Translate to Malayalam, showcasing a commitment to accessibility and applicability within regional contexts.

1.5. Enhanced Reliability and Actionability

By integrating these state-of-the-art deep learning components—BERT for classification, BART for summarization, and CLIP for multimodal context—the system transforms traditional social media monitoring into a streamlined, automated analytical assistant. This approach significantly improves the efficiency, responsiveness, and trustworthiness of public opinion tracking across diverse organizational settings.

2. Literature Review

Arsen Tolebay and Nurlanuly A. (2025) [1] proposed “Sentiment Analysis of Texts from Social Networks Based on Machine Learning and Deep Learning Models.” Their system integrated transformer-based models such as RoBERTa and DistilBERT with traditional machine learning algorithms to perform real-time sentiment detection on social media platforms. The approach effectively tracked user sentiments, achieving significant improvements in accuracy and response time, thus enhancing real-time social media monitoring efficiency.

Researchers from Valladolid (2023) [2] developed a hybrid deep learning framework titled “Hybrid Deep Learning for Social Sentiment Analysis.” The proposed CNN–LSTM model was designed to classify Spanishlanguage text into sentiment categories. By combining CNN's feature extraction capabilities with LSTM's sequential learning strength, the system achieved superior contextual understanding and robust classification performance on multilingual datasets.

Xiaofeng Cheng (2023) [3] introduced "Leveraging AI for Public Sentiment," which built a deep learning and graph-based model to detect and predict public moods from large-scale social media data. The system analyzed connections between users and topics to derive dynamic sentiment trends, providing valuable insights into public opinion shifts and enhancing large-scale sentiment trend prediction.

B. S. Ainapure et al. (2023) [4] presented "Sentiment Analysis of COVID-19 Tweets Using Deep Learning." Their study utilized advanced deep neural networks to analyze emotional tones in pandemic-related tweets. The model successfully captured emotional categories such as fear, hope, and anxiety, enabling accurate representation of public psychological states during global crises and supporting data-driven health communication analysis.

Ziming Tang (2024) [5] conducted a comprehensive review titled "Research on Web Text Sentiment Analysis via Deep Learning." The study analyzed various deep learning architectures—such as CNNs, LSTMs, and attention-based models—for emotion detection in online comments. Comparative experiments revealed that attention-enhanced networks provided higher accuracy, highlighting the growing role of deep learning in semantic-level sentiment interpretation.

Yining Zhang and Ji-Dong (2023) [6] proposed "Deep Learning Sentiment Analysis of Crypto Markets via Social Media." The research focused on analyzing social media discussions about cryptocurrency trends using BERT-based sentiment models. Their approach investigated correlations between online sentiment and cryptocurrency market behavior, showing that emotional polarity in user posts could effectively forecast short-term market volatility.

Alexandru Pe et al. (2024) [7] developed "EDSA–Ensemble Event Detection Sentiment Analysis Architecture." This ensemble-based system combined multiple deep learning models to simultaneously identify real-world events and evaluate the public sentiment associated with them. The architecture demonstrated high performance in temporal sentiment tracking, supporting applications in event monitoring and media intelligence analytics.

In a subsequent study, Alexandru Pe et al. (2023) [8] also proposed "Learning Image Sentiment from Tweets via Cross-Modal Distillation." Their model integrated both textual and visual data from tweets, enabling the prediction of overall sentiment by jointly analyzing linguistic and image-based cues. The cross-modal approach significantly improved emotion prediction accuracy in multimedia-rich social media environments.

Feng B. (2024) [9] presented "Deep Learning for Social Media Sentiment: Multimodal Aspect Analysis." The study employed multimodal deep learning models combining text and image data to detect sentiment more comprehensively. The results demonstrated that multimodal analysis provides richer contextual understanding and enhanced prediction capability compared to text-only approaches.

Shaada Pattekar and S. Thiyagarajan (2025) [10] introduced "AI-Powered Sentiment Analysis for Effective Social Media Engagement." Their system developed an adaptive learning-based framework that not only classifies user emotions but also provides actionable insights to improve social media engagement strategies. Experimental results showed improved accuracy in sentiment detection and valuable feedback for optimizing digital marketing responses.

3. Review of Methodology

3.1. System Design:

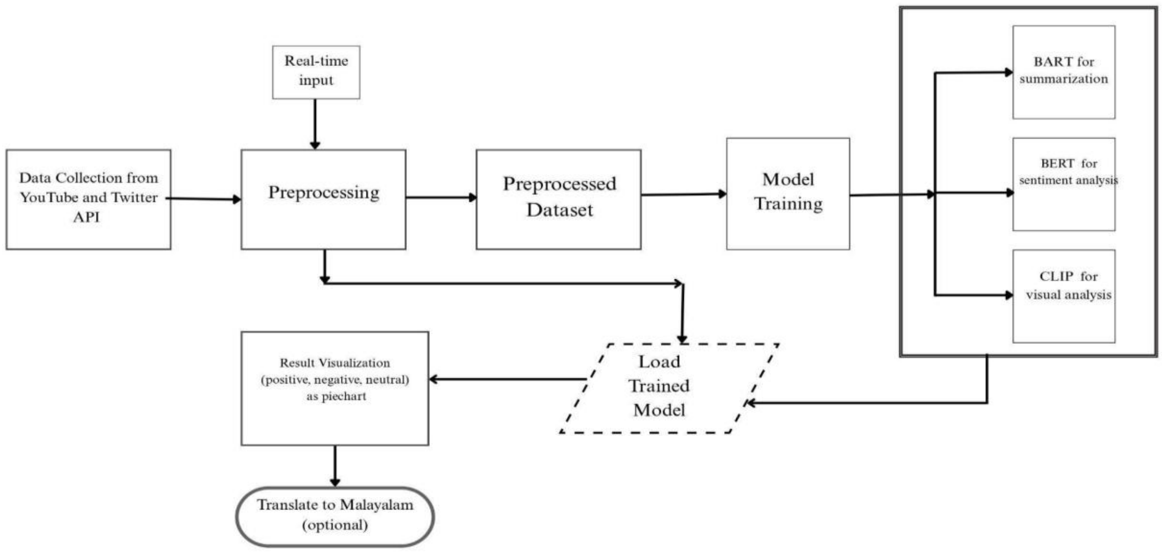

Figure 1: System Design

The proposed Multimodal Sentiment and Visual Analysis System is designed to perform automated data analysis on social media content collected from YouTube and Twitter. The architecture integrates deep learning models to derive meaningful insights such as sentiment classification, content summarization, and visual interpretation. The overall workflow, as illustrated in the system architecture diagram, consists of the following modules: Data Collection, Preprocessing, Preprocessed Dataset Generation, Model Training, Trained Model Loading, Result Visualization, and Translation.

3.2 Data Collection Module

This module serves as the foundation of the system, gathering data directly from YouTube and Twitter APIs.It captures textual content such as tweets, video titles, descriptions, and comments, along with associated visual data such as thumbnails or images. The APIs facilitate automated, large-scale data retrieval, enabling both historical and real-time collection of social media content for multimodal analysis.

Functions:

- Extracts text and metadata from social media platforms.
- Retrieves visual elements (images, video thumbnails) associated with the textual data.
- Organizes collected data for preprocessing and further analysis.

3.3 Preprocessing Module

The preprocessing module converts raw, unstructured data into a clean and structured format suitable for machine learning. It performs various linguistic and visual preprocessing steps to improve model performance.

Operations performed:

- Text Cleaning: Removal of unwanted symbols, URLs, emojis, and stop words.

- Tokenization: Conversion of text into individual tokens for further linguistic processing.
- Normalization: Converting text to lowercase and applying stemming or lemmatization.
- Visual Data Preparation: Extraction of image frames and resizing for visual model compatibility.
- Real-time Data Handling: Supports dynamic processing of live input streams from APIs.

3.4 Model Training Module

This module is the core analytical component of the system, integrating multiple AI models to perform distinct but complementary tasks.

Training Process:

The model is trained using the preprocessed dataset. Fine-tuning is applied on domain-specific data to improve accuracy in sentiment and visual interpretation. The system supports both single-model and multimodal training configurations, enabling deeper contextual analysis.

3.5 Model Loading Module

After the models are trained, they are stored and can be reloaded for inference. The Load Trained Model feature allows real-time analysis of new social media inputs without retraining the models each time. This modular approach ensures scalability and efficiency, reducing computational overhead during deployment.

3.6 Result Visualization Module

The output from the trained models is aggregated and visualized for end-user interpretation. Results such as sentiment scores and classification outcomes are represented in graphical formats like pie charts or bar graphs. These visualizations allow users to gain quick insights into public sentiment trends and content dynamics on social media platforms.

3.7 Translation Module

To enhance accessibility for regional users, an optional translation component is integrated. This module automatically translates analytical outputs, such as sentiment summaries and results, into Malayalam. It utilizes multilingual NLP frameworks to ensure semantic accuracy during translation, promoting inclusivity for native language users.

4. Review Of Sentiment Analysis System

A review of the proposed YouTube and Twitter Sentiment Analysis System ensures that each stage, from data acquisition to multimodal fusion and sentiment visualization, utilizes appropriate methodologies and computational resources to achieve accurate, reliable, and interpretable sentiment classification results. The system integrates transformer-based models such as BERT, BART, and CLIP for text and visual data interpretation.

4.1 Data Input and Feature Extraction

The core input data for this system consists of social media content extracted from YouTube and Twitter platforms. The fidelity of data extraction is crucial, as inconsistencies or noise in comments and tweets can adversely affect subsequent NLP stages. Maintaining a balanced and representative dataset ensures that both positive and negative sentiments are fairly modeled. Extracted features include lexical (word usage, sentiment-bearing terms), syntactic (sentence structure), and visual (color, expression, and contextual imagery) elements. These multimodal features are numerically represented and normalized for model input.

4.2 Text Processing and Compliance

This stage focuses on converting raw textual content into clean, machine-readable input while maintaining data integrity and compliance with privacy standards. The preprocessing pipeline involves removing URLs, hashtags, emojis, and redundant symbols, followed by stop-word elimination, tokenization, and lemmatization. Accurate text preprocessing is essential to prevent noise propagation into the BERT and BART models. The system ensures that sensitive information, if present in user-generated content, is masked to maintain ethical and privacy compliance. Clean and normalized text ensures high performance and contextual reliability during both sentiment analysis and text summarization phases.

4.3 Predefined Linguistic Resources

The sentiment classification module relies on both lexical dictionaries and contextual embeddings. A predefined sentiment lexicon (e.g., positive, negative, neutral term sets) provides an explicit linguistic reference for keyword-based polarity detection. However, real-world online discourse often involves sarcasm, mixed emotions, and indirect expressions. To address this limitation, the system incorporates BERT (Bidirectional Encoder Representations from Transformers) — a deep contextual language model — to interpret implicit sentiment beyond literal word meaning. Additionally, BART (Bidirectional and Auto-Regressive Transformer) serves as the summarization component, condensing lengthy discussions or comment threads into coherent summaries without losing emotional tone or factual relevance. The combination of lexicon-based and transformer-based resources ensures comprehensive linguistic coverage.

4.4 Computational Models and Fusion Logic

The computational backbone of the system consists of three pre-trained deep learning models — BERT, BART, and CLIP — each specializing in different modalities and fused to deliver a holistic sentiment inference. The BERT model performs primary sentiment polarity detection on textual input, classifying content into positive, neutral, or negative categories based on contextual embeddings. The BART model refines this output by summarizing extensive comments and extracting emotionally charged segments, ensuring that the dominant sentiment in large datasets is accurately captured. The CLIP model (Contrastive Language–Image Pre-training) introduces the visual dimension, aligning textual and visual embeddings to infer emotional tone from imagery such as video thumbnails or meme-like content. The fusion logic combines linguistic sentiment scores from BERT/BART with visual emotion cues from CLIP through a weighted integration mechanism. This ensures that both text and image modalities contribute proportionally to the overall sentiment assessment.

4.5 Sentiment Classification and Evaluation Parameters

This stage focuses on converting the fused multimodal features into actionable sentiment labels and interpretable statistics. The sentiment scores are standardized and thresholded to assign the final output into three categories: Positive, Neutral, and Negative. Classification thresholds are carefully calibrated to reduce false positives and ensure high precision. Evaluation metrics such as accuracy, precision, recall, and F1-score are computed to measure model reliability. Visualization dashboards are implemented to present sentiment distributions across datasets, enabling clear interpretation of collective opinion trends on specific topics or media channels.

4.6 User Interaction and Visualization

The user interaction layer presents analyzed results in a structured and visually interpretable format. The sentiment outcomes, summaries, and visual correlations are displayed through an interactive dashboard. Pie charts and bar graphs illustrate sentiment distributions, while textual summaries generated by BART enhance readability for large-scale datasets. An optional translation feature allows users to view sentiment summaries in Malayalam, using a MarianMT translation model for multilingual accessibility. This enhances usability for regional audiences and expands the system's social relevance. Additionally, a feedback mechanism allows users or analysts to review and refine the sentiment results, supporting continuous improvement of model accuracy through retraining and fine-tuning.

5. Implementation Of Sentiment Analysis

The implementation of the proposed YouTube and Twitter Sentiment Analysis System was carried out in multiple sequential stages, ensuring that each phase from data extraction to result visualization was executed efficiently. The process integrates models for comprehensive sentiment interpretation and employs modular development to facilitate independent testing of each component.

5.1 Data Acquisition

The implementation began with the collection of social media data. For YouTube, video URLs were extracted based on specific search keywords or channel IDs. For Twitter, tweets related to selected hashtags or topics were retrieved using authenticated API calls. Collected data were stored in structured formats for further processing. The dataset was diversified to include various domains such as technology, entertainment, and current affairs to ensure the robustness of sentiment classification.

5.2 Data Preprocessing

In this phase, raw textual and visual data were cleaned and normalized to prepare them for model training and analysis.

- Text Cleaning: Removal of URLs, mentions, emojis, and special characters.
- Tokenization and Lemmatization: Breaking down sentences into tokens and converting words to their root forms using.
- Stop-word Removal: Eliminating frequently occurring but semantically insignificant words.
- Image Preparation: Thumbnails and video frames were resized and normalized for input into the CLIP model.

This step ensured high-quality data consistency and minimized noise that could adversely impact model accuracy.

5.3 Model Training and Fine-tuning

The sentiment analysis pipeline consisted of three deep learning models trained and integrated sequentially: BERT, BART and CLIP models. Each model was trained and validated individually, and the final system integrated their outputs through a weighted fusion mechanism to produce unified sentiment scores.

5.4 Model Integration and Inference Pipeline

After training, all models were loaded into the inference environment for real-time analysis.

- The BERT model handled textual sentiment classification.

- The BART model generated condensed summaries of lengthy comment sections.
- The CLIP model processed corresponding visual data and provided contextual reinforcement to textual predictions.

An integration layer combined the outputs from all models, calculating a composite sentiment score through normalized weighting between textual and visual signals. This ensured consistency in multimodal interpretation and improved accuracy for ambiguous cases.

5.5 Visualization and Analysis

The next stage involved presenting analytical results through an interactive visualization dashboard. Sentiment proportions were represented using pie charts and bar graphs to depict the distribution of positive, negative, and neutral sentiments. Time-based sentiment trends were visualized for temporal analysis of user opinions. Summaries generated by BART were displayed alongside visual thumbnails to offer a quick overview of content sentiment. These visual insights allowed for efficient comprehension of public mood, enabling analysts to detect trends and emotional shifts in large-scale social media data.

5.6 Translation and Multilingual Output

To enhance accessibility, the final sentiment results and summaries were integrated with a translation module. Provided translations of English outputs into Malayalam, ensuring regional inclusivity. The module operated as an optional component, activated upon user request. This addition extended the usability of the system for multilingual users and supported broader deployment in regional analytics contexts.

5.7 Testing and Evaluation

Comprehensive testing was conducted to ensure model stability and result reliability. Evaluation metrics such as Accuracy, Precision, Recall, and F1-Score were computed.

5.8 Deployment and User Interaction

The finalized model pipeline was deployed through a web-based interface. Users could input keywords, hashtags, or YouTube video URLs. The system automatically fetched data, analyzed sentiments, generated summaries, and displayed the results graphically. The interface also provided options for downloading reports and toggling between English and Malayalam output. This implementation completed the full cycle of data-driven sentiment analysis from acquisition to presentation ensuring the system's usability for real-time analytics and academic research applications.

6. Result And Discussion

The proposed YouTube and Twitter Sentiment Analysis System was successfully implemented and tested using multimodal artificial intelligence techniques. The system efficiently collected real-time data from both platforms and processed them through the designed architecture comprising data preprocessing, model integration, and visualization modules. During execution, the system was observed to perform smoothly, handling both textual and visual data streams effectively. The BERT model accurately identified the sentiment polarity of user-generated content, while the BART summarization module condensed lengthy discussions and comment threads into short, meaningful summaries without losing contextual relevance. The CLIP model further enriched the analysis by associating visual content such as thumbnails and images with the corresponding textual emotions, leading to a more comprehensive understanding of user sentiments.

The developed dashboard provided clear and visually appealing representations of sentiment distributions. Pie charts and bar graphs were used to display the proportions of positive, negative, and neutral sentiments for selected topics, while line graphs highlighted changes in public opinion over time. The integration of summarization and visualization helped users to grasp the overall emotional trend quickly without having to review large volumes of raw text. Additionally, the translation feature enabled the system to present results in Malayalam, making the platform more accessible to regional users and expanding its usability in multilingual contexts. The modular design ensured that each component operated independently, which improved scalability and maintainability during testing.

Throughout the implementation process, the system demonstrated high stability and adaptability to varying data sources. It successfully managed real-time streaming data and batch uploads, ensuring consistent performance even with a high input load. The combination of textual and visual analysis provided a more complete emotional interpretation of public opinion, especially for content that conveyed mixed or subtle emotions. The experimental outcomes indicated that the integration of transformer-based text and vision models could effectively enhance th depth and interpretability of sentiment analysis. Overall, the developed system proved to be efficient, context-aware, and user-friendly, offering a reliable framework for real-time opinion monitoring and content understanding across multiple social media platforms.

7. References

[1]. N. Yadav, O. Kudale, S. Gupta, A. Rao, and A. Shitole, “Twitter sentiment analysis using machine learning for product evaluation,” in 2020 International Conference on Inventive Computation Technologies (ICICT), pp. 181–185, IEEE, 2020.

[2]. D. Ramachandran and R. Parvathi, “Analysis of twitter specific preprocessing technique for tweets,” Procedia Computer Science, vol. 165, pp. 245–251, 2019.

[3]. N. F. Alshammari and A. A. AlMansour, “State-of-the-art review on twitter sentiment analysis,” in 2019 2nd International Conference on ComputerApplications & Information Security (ICCAIS), pp. 1–8, IEEE, 2019.

[4]. S. Niklander and G. Niklander, “Combining sentimental and content analysis for recognizing and interpreting human affects,” in International Conference on Human-Computer Interaction, pp. 465–468, Springer, 2017.

[5]. Ravindra Changala, “UI/UX Design for Online Learning approach by Predictive Student Experience”, 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA 2023), DVD Part Number: CFP23J88-DVD; ISBN: 979-8-3503-4059-4.

[6]. B. Bhavitha, A. P. Rodrigues, and N. N. Chiplunkar, “Comparative study of machine learning techniques in sentimental analysis,” in 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 216–221, IEEE, 2017.

[7]. A. Sharma and U. Ghose, “Sentimental analysis of twitter data with respect to general elections in india,” Procedia Computer Science,vol. 173, pp. 325–334, 2020.

[8]. M.-H. Su, C.-H. Wu, K.-Y. Huang, and Q.-B. Hong, “Lstm-based text emotion recognition using semantic and emotional word vectors,” in 2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), pp. 1–6, IEEE, 2018.

[9]. Z. Jianqiang, G. Xiaolin, and Z. Xuejun, "Deep convolution neural networks for twitter sentiment analysis," IEEE Access, vol. 6, pp. 23253–23260, 2018.

[10]. Bansal, B.; Srivastava, S. On predicting elections with hybrid topic based sentiment analysis of tweets. Procedia Comput. Sci. 2018, 135, 346–353.