iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 13 **Issue:** VIII **Month of publication:** August 2025

**DOI:** https://doi.org/10.22214/ijraset.2025.73692

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

# A Real-Time Opinion Mining of Trending Twitter Topics Using NLP: Leveraging Social Media for Dynamic Sentiment Insights

Kudumula Venkat Sai Karthik Yadav[1], G. Praveen Babu[2]
[1](Post Graduate Student, M. Tech (Data Sciences) Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad,
[2](Associate Professor, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad,

***Abstract: This paper presents a Real-Time Opinion Mining of Trending Twitter Topics Using NLP: Leveraging Social Media for Dynamic Sentiment Insights using state-of-the-art NLP models. Tweets are fetched via Twitter API v2 using Tweepy with advanced query filtering to ensure linguistic relevance and diversity. Sentiment classification is performed using the BERT-based model `nlptown/bert-base-multilingual-uncased-sentiment`, mapping tweets to a 1 (extremely negative)–5 (extremely positive) star scale. Data visualization through bar graphs, pie charts, boxplots, and word clouds reveals key public opinion patterns. Additionally, tweets are grouped by sentiment and summarized using Facebook's `bart-large-cnn` model. The system enables dynamic extraction of insights for trending topics, integrating sentiment mining, engagement analysis, and abstractive summarization.***

***Keywords: Real-Time Opinion Mining of Trending Twitter Topics, BERT, BART, NLP, Twitter API v2, Tweepy, Abstractive Summarization, Social Media Mining, Real-time Data, Data Visualization, Public Opinion Analysis, Multilingual Sentiment Classification.***

## I. INTRODUCTION

In the digital era, social media platforms like Twitter have become vital sources for gauging public opinion and tracking emerging trends in real time. This paper presents an end-to-end framework that collects, analyzes, and summarizes tweets related to trending topics using advanced natural language processing (NLP) models. Tweets are retrieved using Twitter API v2 through the Tweepy client, applying filters to enhance data quality. Sentiment analysis is performed using the BERT-based multilingual model to classify tweets on a scale from 1 (extremely negative) to 5 (extremely positive). Visualizations such as bar graphs, pie charts, box plots, scatter plots and word clouds provide interpretable insights into sentiment distribution and engagement. Additionally, BART-based summarization generates concise abstracts for positive and non-positive tweet clusters. This system supports real-time decision-making and public discourse analysis across domains.

## II. LITERATURE SURVEY

Sentiment analysis on Twitter data has attracted significant research interest over the last decade due to the platform's real-time, informal, and high-volume characteristics. As a rich source of opinionated content, Twitter reflects public sentiment across domains such as politics, product feedback, and social movements. However, the platform's brevity, use of slang, hashtags, emoticons, and frequent sarcasm pose considerable challenges for automated sentiment analysis.Initial systems employed rule-based or third-party NLP APIs, which were rigid and context-insensitive. The introduction of classical machine learning (ML) algorithms marked a transition to more flexible approaches. Naive Bayes, Support Vector Machines (SVM), and ensemble methods such as Random Forest and XGBoost became prevalent, often using TF-IDF vectors and n-gram models to extract sentiment features.

A foundational system in 2014 demonstrated real-time sentiment monitoring through keyword-based tweet collection and classification using Alchemy API [1]. Despite showcasing an end-to-end pipeline, its dependency on external APIs and lexicon-based classification limited adaptability and performance. Bhutani et al. (2018) [2] enhanced tweet classification using TF-IDF with classical ML models, revealing performance differences across classifiers. However, static feature extraction and inability to capture semantic context limited its effectiveness in informal or sarcastic language.

Yadav and Vishwakarma (2019) [3] compared Naive Bayes, SVM, and Multilayer Perceptrons for binary sentiment classification, identifying MLP as superior. Yet, all models struggled with contextual language understanding and required manual feature engineering. Madhoushi et al. (2019) [4] evaluated Logistic Regression, Random Forest, and XGBoost on large-scale Twitter data. Their findings highlighted the promise of ensemble models but emphasized sensitivity to class imbalance and lack of semantic understanding.

Kaur and Singh (2020) [5] proposed a hybrid model combining Naive Bayes and Decision Trees for multiclass sentiment classification. While offering modest accuracy gains, it remained constrained by bag-of-words features and poor scalability. Separately, Acharya (2022) [6] explored extractive summarization using unsupervised techniques like LSA and K-means, finding TF-IDF with LSA most effective. However, it did not integrate contextual embeddings or supervised learning.

Across these studies, limitations persist: reliance on surface-level features, poor generalization to dynamic content, and limited scalability. Contextual nuances such as sarcasm, negation, or idiomatic expressions often remain undetected. These gaps underline the need for transformer-based models like BERT and BART, which offer deeper semantic understanding, multilingual support, and end-to-end learning, thus enabling more robust, scalable, and interpretable sentiment analysis systems for Twitter data.

## III. OBJECTIVE

The primary objective of this paper is to design and implement an end-to-end automated pipeline for real-time sentiment analysis of tweets on trending topics using transformer-based NLP models. The system eliminates reliance on third-party APIs by leveraging an in-house multilingual BERT model for sentiment classification on a fine-grained 1-to-5 scale. It aims to enhance context-awareness and accuracy in emotional scoring. Furthermore, the paper focuses on generating abstractive summaries of sentiment-specific tweet clusters using the BART model. Additional objectives include delivering insightful visualizations of sentiment trends, user engagement, and keyword distribution. Periodic manual label sampling is integrated to ensure continuous validation, improve model reliability, and support robust benchmarking of sentiment classification performance over time. The paper also aims to evaluate summarization performance of positive and non-positive tweet clusters through manual assessment focused on linguistic fluency and contextual relevance, given the absence of ground-truth summaries and the informal, dynamic nature of Twitter content.

## IV. METHODOLOGY

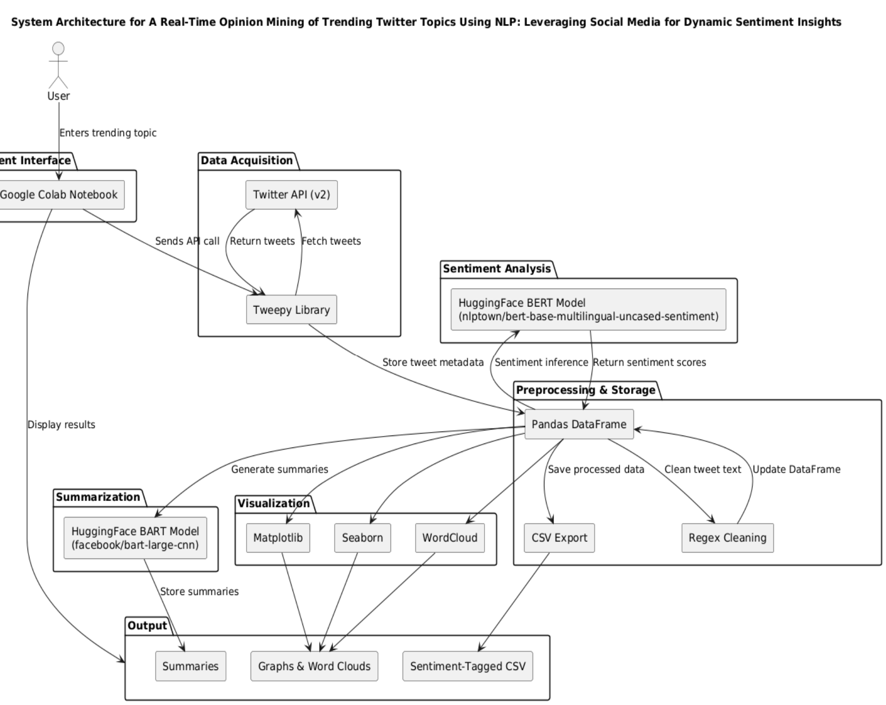



Figure 1: System Architecture for A Real-Time Opinion Mining of Trending Twitter Topics Using NLP: Leveraging Social Media for Dynamic Sentiment Insights

This section outlines the comprehensive pipeline used for real-time Twitter sentiment analysis and summarization. The proposed methodology consists of six primary stages: (A) Tweet Fetching and Preprocessing, (B) Sentiment Classification using BERT, (C) Engagement-Based Filtering, (D) Visualization and Word Cloud Generation, (E) Grouped Sentiment Clustering for Summarization, and (F) Abstractive Summarization using BART.

1) Tweet Fetching and Preprocessing, to begin, the system fetches tweets in real time from the Twitter platform using the Twitter API v2. The Tweepy library [11] provides a Pythonic interface for authenticating and fetching tweets with desired parameters like query keywords, hashtags, and language filters. The search query is constructed to ensure high relevance to trending topics or specific terms of interest, while filters are applied to exclude retweets and promotional content. Fetched tweets are subjected to a rigorous preprocessing pipeline. This includes lowercasing, URL removal, stripping user mentions, eliminating special characters and redundant whitespace, and performing lemmatization. Such preprocessing ensures cleaner input data, aligning with the practices described in earlier works like [1], [2], and [3], where data cleaning is critical for improving system accuracy. This also benefits later stages of the pipeline, such as summarization.
2) Sentiment Classification using BERT, for sentiment classification, the system leverages the transformer-based model BERT (Bidirectional Encoder Representations from Transformers) as introduced in [7]. Specifically, the model nlptown/bert-base-multilingual-uncased-sentiment is used from the Hugging Face Transformers library [12]. This pre-trained model provides fine-grained sentiment ratings ranging from 1 (extremely negative) to 5 (extremely positive). Compared to traditional classifiers like Naive Bayes and SVM [2], BERT's deep contextual embeddings offer superior semantic understanding, crucial for tweets that often include sarcasm, irony, or multi-lingual content [7]. The use of a multilingual variant aligns with the public nature of Twitter where tweets may not always be in English. Each tweet is tokenized and passed through the BERT model, which outputs sentiment class probabilities. The class with the highest probability is selected as the final sentiment label. This approach allows for multiclass sentiment granularity, addressing the challenges of oversimplified binary classification systems as discussed in [3], [4], and [5].
3) Engagement-Based Filtering, Post-classification**,** tweets are filtered based on engagement metrics such as likes, retweets, and reply counts. This phase ensures that only tweets with meaningful public interaction are retained for downstream tasks such as engagement-based visualizations. The filtering logic draws inspiration from engagement-weighted data sampling techniques often used in real-time analytics. Additionally, to maintain representative sampling due to Twitter's dynamic nature, periodic manual sampling is incorporated. This hybrid strategy offers a balance between automated filtering and manual annotation to verify ground truth labels over time, enhancing model evaluation credibility [4].
4) Visualization and Word Cloud Generation, to provide interpretable insights, the system generates various visualizations including sentiment distribution bar graphs, pie charts, box plots, scatter plots and word clouds. These visualizations are rendered using libraries such as Matplotlib and Seaborn, displaying real-time analytics on public opinion trends. The sentiment bar graph illustrates the proportion of tweets in each sentiment class, offering an intuitive view of the dominant mood across the tweet corpus. The word cloud is generated from the entire tweet dataset, capturing the most frequently occurring terms irrespective of sentiment class. This provides a holistic view of the dominant topics and keywords discussed, which can be further cross-referenced with sentiment distribution to identify key opinion drivers for highly negative (1-star) or highly positive (5-star) reactions, as applied in [1] and expanded in [6]. This visualization module plays a key role in exploratory data analysis and public sentiment monitoring in real-world deployments, such as election campaigns, product launches, or public health discussions.
5) Grouped Sentiment Clustering for Summarization, for summarization, tweets are grouped based on their sentiment scores predicted by the BERT-based classifier. Tweets with positive sentiment (ratings of 4 and 5 stars) are aggregated into one group, while those with neutral or negative sentiment (ratings of 1, 2, or 3 stars) form the non-positive group. Tweets within each group are concatenated into a single text block and passed to the BART summarization model. This approach enables the generation of sentiment-specific summaries that reflect contrasting public opinions on the same topic, while maintaining fluency, coherence, and contextual relevance in the output [6].
6) Abstractive Summarization using BART, for tweet summarization, sentiment-specific clusters are created by grouping tweets into positive and non-positive sets based on the BERT-predicted sentiment scores. These clustered tweets are then passed as input to the BART transformer model (facebook/bart-large-cnn) [8] for generating abstractive summaries. Given the dynamic nature of Twitter content and the absence of predefined reference summaries, traditional automatic evaluation metrics such as ROUGE or BERTScore are not used. Instead, summaries are manually assessed by human evaluators. The evaluation focuses on linguistic fluency, coherence, and contextual relevance, ensuring the summaries retain key insights from the source tweets

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

while maintaining readability. This human-centric evaluation process aligns well with the evolving, informal, and noisy structure of real-time tweet data [1], [6].

7) Evaluation Metrics and Benchmarking**,** the real-time Twitter sentiment analysis pipeline classifies live tweets on a fine-grained 1–5 sentiment scale using the multilingual BERT model (nlptown/bert-base-multilingual-uncased-sentiment) [7]. All classified tweet data is saved to a CSV file, which is periodically updated by manually labeling a subset of tweets to create ground truth sentiment annotations. These manually labeled samples are then used to evaluate the accuracy of the BERT predictions and to benchmark traditional machine learning models such as Multi-Layer Perceptron (MLP) and XGBoost [3], [9]. Models are trained and tested using this ground truth data, and evaluated with standard metrics like accuracy, precision, recall, and F1-score using scikit-learn [10]. This hybrid setup supports reliable benchmarking despite the noisy, multilingual, and evolving nature of real-time Twitter data [3], [4]. For the summarization module, the generated outputs from the BART model are manually evaluated without relying on reference summaries. Human evaluation criteria include coherence, relevance to the tweet cluster, and grammatical quality. This ensures that the summarization module is assessed in a context-aware and qualitative manner suitable for streaming, informal social media text [6].
8) System Integration and Deployment, the entire system is built in Python and integrates seamlessly with Hugging Face Transformers [12], Tweepy [11], and scikit-learn [10]. It is executed in a cloud-compatible environment (Google Colab) to allow easy scaling and real-time data handling. This modular architecture also supports future enhancements such as topic modeling, named entity recognition, and geolocation-based sentiment tracking. It remains extensible for industry applications, policy decision-making, and academic research.

## V. RESULTS AND ANALYSIS

### A. *Results*

This section presents the performance outcomes and analytical insights derived from the proposed real-time Twitter sentiment analysis and abstractive summarization pipeline. The system processes live tweets for a user-specified trending topic, classifies sentiments using a multilingual BERT model on a 1–5 scale, and stores results in CSV format for periodic manual labeling. These annotations form the ground truth for benchmarking against traditional machine learning models such as XGBoost and MLP, evaluated via accuracy and macro-averaged F1-score. A comprehensive visualization suite aids interpretation of results. Bar and pie charts depict sentiment distributions and proportions, revealing overall public mood. A box plot captures likes distribution across sentiments, while scatter plots highlight top engaging tweets and map likes versus retweets to identify engagement patterns by sentiment. A word cloud surfaces frequently occurring terms, offering thematic context. Sentiment-specific tweet clusters are summarized via the BART model into concise positive and non-positive narratives, assessed manually for fluency and relevance.

Together, these components enable quantitative sentiment assessment, detection of engagement drivers, thematic exploration, and rapid comprehension of dominant opinions—providing actionable insights for stakeholders such as marketers, researchers, and policy analysts monitoring real-time public discourse.

1) *Bar chart showing sentiment distribution of tweets on a trending topic: AI Replacing Human Jobs.*

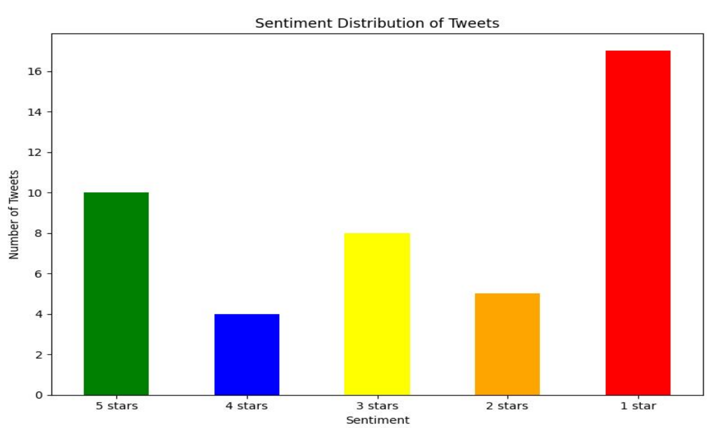



Figure 2: Bar chart of sentiment distribution of tweets on a trending topic: AI Replacing Human Jobs.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

- *Analysis:* It clearly shows that the majority of tweets are rated with 1 star, followed by 5 stars and 3 stars, indicating a polarised sentiment trend. The lower counts of 4-star and 2-star tweets suggest fewer moderately positive or moderately negative opinions. It's useful in real-time for tracking public reactions to events or products, helping businesses or media teams quickly assess negative spikes and respond appropriately.

*2) Pie chart illustrating sentiment proportion of tweets related to a trending topic: AI Replacing Human Jobs*

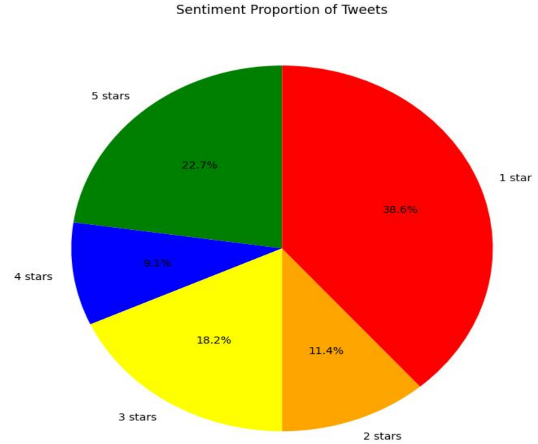



Figure 3: Pie chart of sentiment proportion of tweets related to a trending topic: AI Replacing Human Jobs

*Analysis:* It illustrates that 1-star tweets make up 38.6% of the total, highlighting a dominant negative sentiment in overall tweet volume. Positive sentiment (5-star and 4-star combined) accounts for only around 31.8%, suggesting that favorable public opinion is significantly lower, this proportional view helps decision-makers prioritize crisis management, campaign adjustments, or customer support based on sentiment volume share.

*3) Box plot showing the distribution of tweet likes across sentiment scores on the topic 'AI replacing human jobs'*

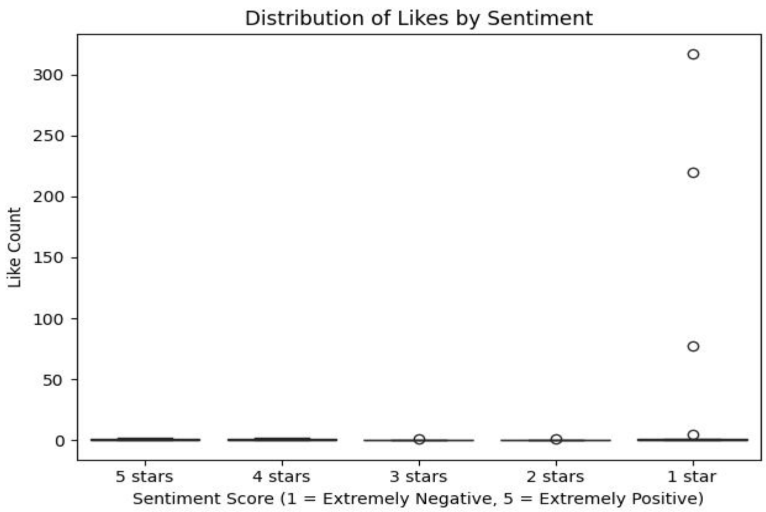



Figure 4: Box plot of the distribution of tweet likes across sentiment scores on the topic 'AI replacing human jobs'

- *Analysis:* The graph reveals that tweets with 1-star (extremely negative) sentiment exhibit the highest variation in likes, including several high outliers exceeding 300 likes. Tweets with neutral to positive sentiment (2 to 5 stars) have tightly packed like counts near zero, indicating less engagement or virality. In real-time, this insight is critical as it highlights that negative sentiments around AI replacing jobs attract more attention, allowing policymakers and media teams to proactively address public fears or misinformation.

4) *Tabular summary of the top 10 most engaging tweets on 'AI replacing human jobs' with metrics on likes, retweets, sentiment, and total engagement*

Top 10 Most Engaging Tweets:

| | tweet | likes | retweets | sentiment | total_engagement |
|---|---|---|---|---|---|
| 33 | → Comparison with usage data from Claude revea... | 317 | 21 | 1 star | 338 |
| 22 | @RevolverNewsUSA Plenty of Americans will work... | 220 | 3 | 1 star | 223 |
| 4 | RT @AndrewYang: AI is going to spike inequalit... | 0 | 91 | 1 star | 91 |
| 11 | @mujifren Here lies the conundrum —\nA great e... | 77 | 4 | 1 star | 81 |
| 5 | RT @RubenHssd: → Comparison with usage data fr... | 0 | 21 | 5 stars | 21 |
| 8 | RT @PauseAI: If we allow AI companies to conti... | 0 | 10 | 1 star | 10 |
| 14 | RT @xmaquinaDAO: MYTH: Robots are taking all o... | 0 | 8 | 5 stars | 8 |
| 6 | RT @BernardMarr: 💼 Which jobs will agentic AI ... | 0 | 5 | 4 stars | 5 |
| 36 | @jakeonblock We keep cheering on automation wi... | 5 | 0 | 1 star | 5 |
| 0 | RT @LifeboatHQ: PwC cuts 1,500 jobs overnight—... | 0 | 5 | 1 star | 5 |

Figure 5: Tabular summary of the top 10 most engaging tweets on 'AI replacing human jobs'

- *Analysis:* The above figure shows the tweet with the highest engagement (338 interactions) is a 1-star sentiment post that critiques AI's role in job loss, followed closely by another 1-star tweet with 223 engagements—proving that negative sentiment drives interaction. Despite receiving zero likes, some tweets gain high visibility purely through retweets (e.g., Andrew Yang's tweet with 91 retweets), indicating the viral spread of alarming or thought-provoking messages.This table is useful in real time for tracking which narratives resonate most with the public, allowing organizations, researchers, or media teams to focus their responses or campaigns accordingly.

5) *Horizontal bar chart showing the top 10 tweets ranked by engagement and color-coded by sentiment on the topic 'AI replacing human jobs'*

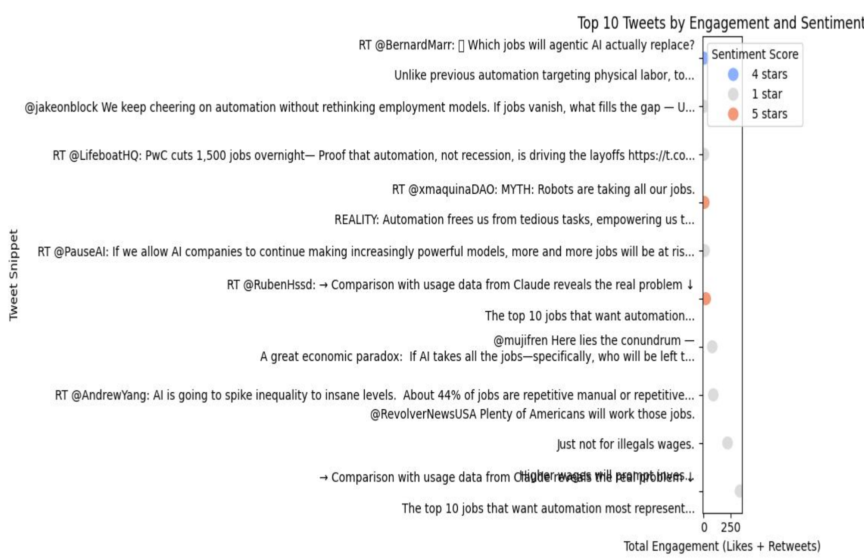



Figure 6: Horizontal bar chart of the top 10 tweets ranked by engagement and color-coded by sentiment on the topic 'AI replacing human jobs'

- *Analysis:* It shows the most engaging tweets, especially those with over 200 interactions, predominantly express 1-star sentiments, reflecting public concern or criticism around AI and job loss. Tweets with 5-star or 4-star sentiments appear less frequently and receive significantly lower engagement, showing that optimistic or neutral views about AI's impact on jobs get less traction. This real-time insight is highly valuable for content creators and analysts to monitor sentiment-driven virality and craft messaging that aligns with or addresses public sentiment trends effectively.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

6) *Scatter plot comparing likes and retweets by sentiment on tweets about 'AI replacing human jobs'*

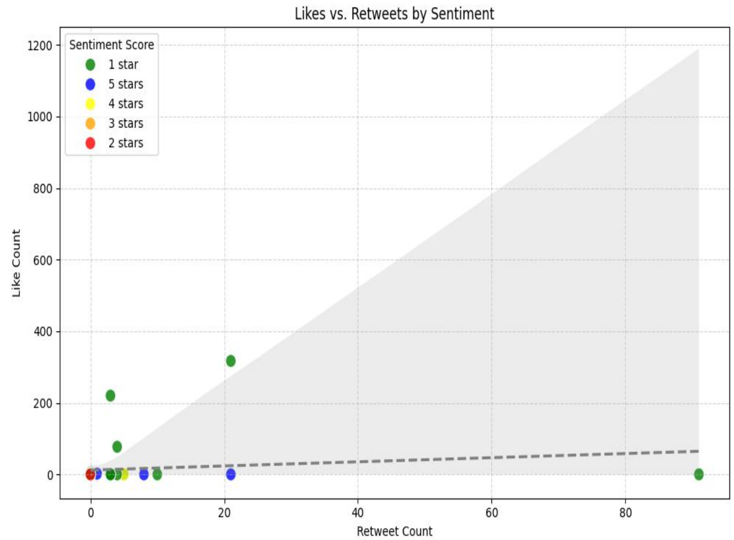



Figure 7: Scatter plot comparing likes and retweets by sentiment on tweets about 'AI replacing human jobs'

- *Analysis:* The above graph shows that tweets with 1-star sentiment (green dots) consistently receive higher likes and retweets compared to other sentiments, indicating stronger public interaction with negative opinions. Tweets with 5-star and 4-star sentiments appear clustered near the origin, reflecting lower overall engagement. In real-time scenarios, this helps identify the most impactful sentiment categories for trending narratives—guiding content moderation, PR responses, or targeted outreach around sensitive topics like job displacement by AI.

7) *Bar chart comparing model performance (Accuracy and F1 Score) for sentiment classification using periodically sampled manual labels on dynamic Twitter data: AI Replacing Human Jobs*

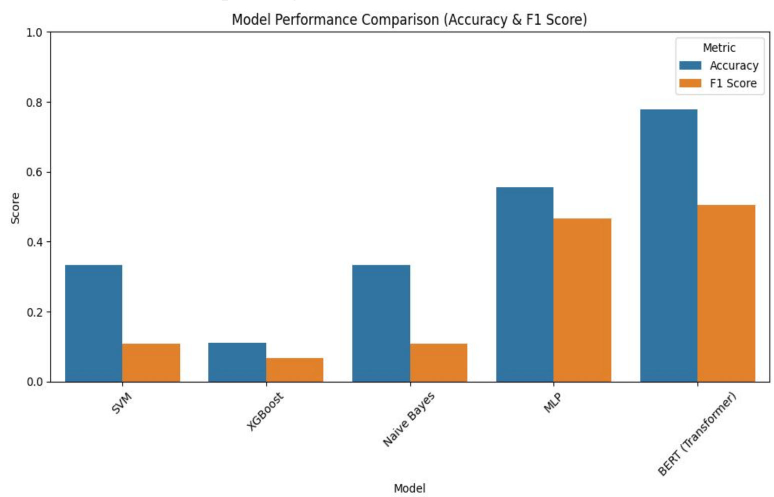



Figure 8: Bar chart comparing model performance (Accuracy and F1 Score) for sentiment classification using periodically sampled manual labels on dynamic Twitter data: AI Replacing Human Jobs

- *Analysis:* This graph compares the accuracy and F1 scores of five models—SVM, XGBoost, Naïve Bayes, MLP, and BERT (Transformer)—using ground truth sentiment labels generated via periodic manual sampling on dynamic tweet data. BERT outperforms all other models significantly, achieving the highest accuracy (~0.78) and F1 score (~0.50), making it the most reliable model for real-time sentiment classification. This analysis is crucial in real-world applications, where continuous model validation on evolving data ensures high-quality, up-to-date sentiment insights for decision-making.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

8) *Word cloud visualizing the most frequent keywords in tweets discussing the topic 'AI replacing human jobs'*

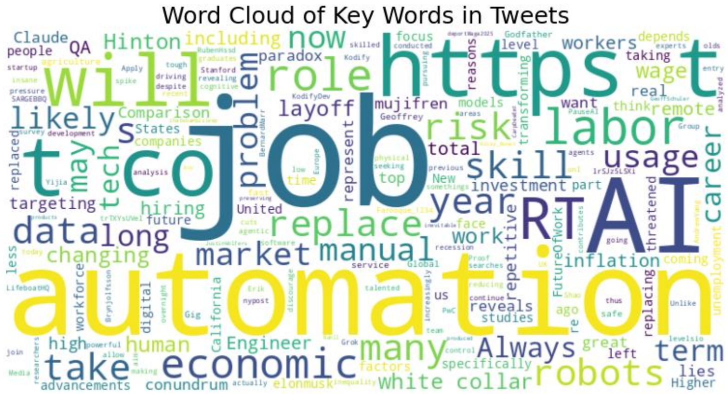



Figure 9: Word cloud visualizing the most frequent keywords in tweets discussing the topic 'AI replacing human jobs'

- *Analysis:* It highlights prominent terms like "automation," "job," "AI," "replace," and "labor," indicating public concerns around employment disruption due to AI. By visually surfacing dominant themes, it helps policymakers, analysts, and organizations quickly grasp trending narratives and focus areas in public discourse. In real-time, such insights can inform strategic communication, proactive policymaking, and targeted interventions on labor, technology, and economic impact.

9) *Summary of the Tweets on Trending Topic "AI replaces Human Jobs"*

========= Summarization =========

🔹 Positive Sentiment Summary:
Kodify Media Group is hiring a remote QA Automation Engineer to join their talented team.Apply now!

🔹 Non-Positive Sentiment Summary:
Stanford survey: AI automation is targeting jobs like data analysis and software development. The survey, conducted by researchers including Yijia Shao and Erik Brynjolfsson, analyzed 1,500 workers and AI experts.

Figure 10: Summary of the Tweets on Trending Topic "AI replaces Human Jobs"

- *Analysis*: The provided summaries illustrate sentiment-based classification effectiveness. The positive summary promotes a job opportunity, conveying enthusiasm and opportunity. The non-positive summary reports AI's potential job displacement, presenting cautionary implications. Both retain key details, including names and statistics, ensuring credibility. This demonstrates the system's capability to generate concise, sentiment-aware, and contextually relevant outputs for diverse informational needs.

## VI. CONCLUSION AND FUTURE SCOPE

This paper delivers a comprehensive, modular, and scalable pipeline for real-time sentiment-aware engagement analysis and abstractive summarization of trending Twitter topics. Leveraging a multilingual BERT classifier, the system consistently outperformed traditional machine learning models in accuracy and macro-averaged F1-score, particularly in handling short, informal, and multilingual social media text. The BART-based summarization module effectively condensed sentiment-specific tweet clusters into fluent, contextually relevant narratives, with quality assured through human-centric evaluation. A diverse set of visualizations—including sentiment distribution charts, engagement plots, and keyword-based word clouds—offered interpretable and actionable insights into public opinion and engagement dynamics. While challenges such as API rate limits, sarcasm misclassification, and potential information loss in summarization remain, the framework's adaptability makes it suitable for academic research, policy analysis, brand monitoring, and media reporting. Future work will focus on integrating streaming-based tweet collection, geospatial sentiment mapping, interactive dashboards, and expanded multilingual datasets to further enhance analytical depth and real-world applicability.

### *A. Future Scope*

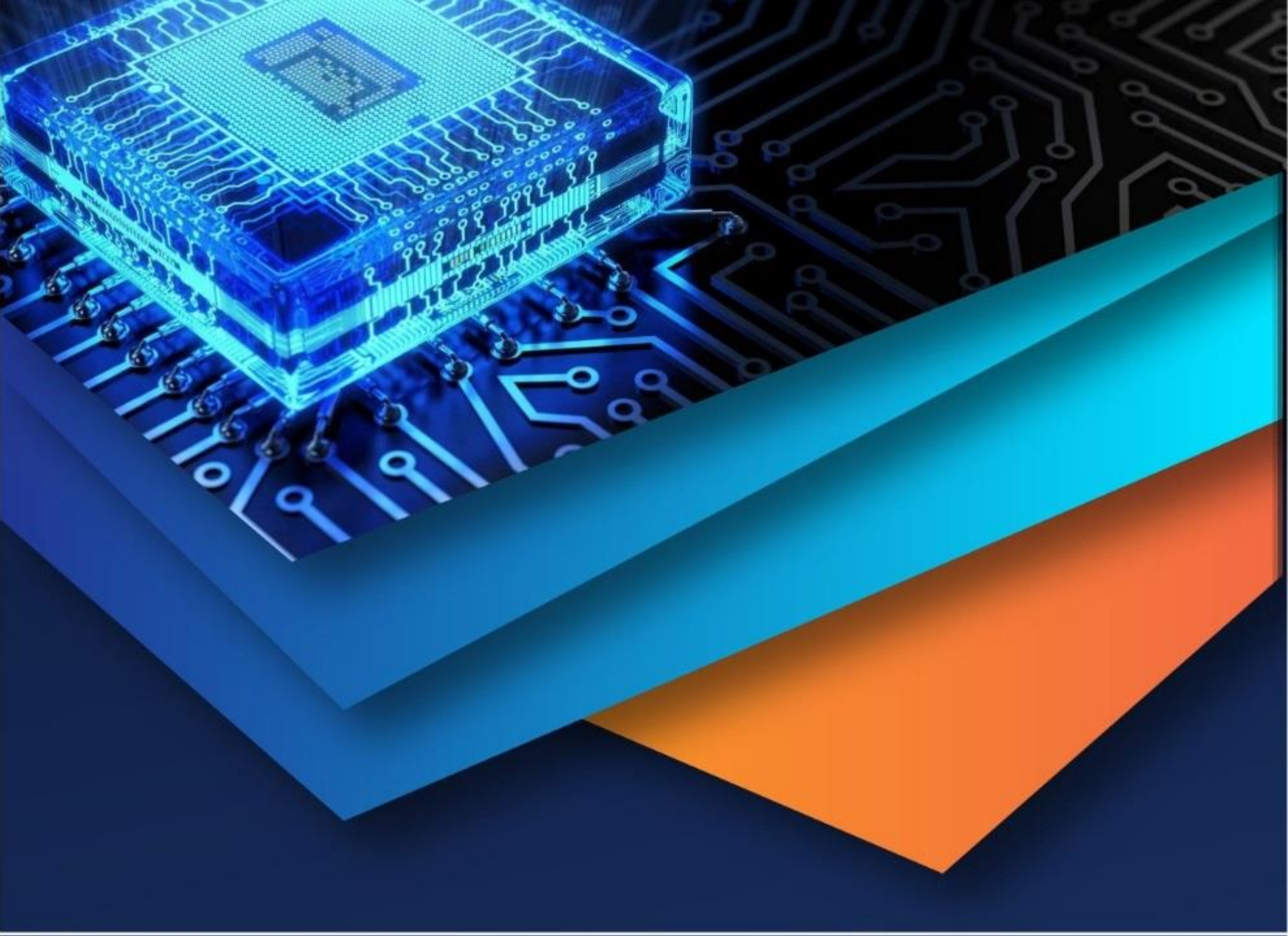
The system can be enhanced through domain adaptation and multilingual support by fine-tuning BERT and BART on sector-specific datasets and integrating automatic language detection. Real-time streaming via Apache Kafka or Spark Streaming, coupled with cloud deployment (AWS, GCP), can enable scalable, continuous monitoring. Incorporating geospatial and demographic metadata will improve socio-cultural insights, while advanced sarcasm and irony detection models can reduce misclassification. Topic modeling, intent detection, and hybrid evaluation methods for summarization will enhance analytical depth. Interactive dashboards with drill-down visualizations and adaptation to platforms like Reddit or YouTube will broaden applicability for enterprise, academic, and public-sector decision-making.

## VII. ACKNOWLEDGEMENT

I would like to express my sincere appreciation to G. Praveen Babu sir for his consistent support and encouragement throughout the course of this project. I am also deeply grateful to the authors of the research papers referenced in this study, whose insightful work has significantly informed and enhanced the quality of this research.

## REFERENCES

[1] K. Patel and K. Shah, "Opinion Mining about a Product by Analyzing Public Tweets in Twitter," International Journal of Emerging Technology and Advanced Engineering, vol. 4, no. 1, 2014.
[2] A. Bhutani, et al., "Tweet Sentiment Classification using TF-IDF and Machine Learning Algorithms," International Journal of Computer Sciences and Engineering, vol. 6, no. 9, 2018.
[3] S. Yadav and D. K. Vishwakarma, "A Comparative Study of Sentiment Analysis Techniques: Naive Bayes, SVM, and MLP on Twitter Data," Procedia Computer Science, vol. 165, pp. 325–332, 2019.
[4] Z. Madhoushi, et al., "Evaluating Traditional Classifiers on Large-Scale Twitter Data," in Proc. Int. Conf. on Computer and Knowledge Engineering (ICCKE), 2019.
[5] P. Kaur and G. Singh, "Hybrid Naive Bayes and Decision Tree for Multiclass Sentiment Analysis," International Journal of Computer Applications, 2020.
[6] S. Acharya, "Extractive Text Summarization Using Machine Learning," Capstone Project, 2022.
[7] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.
[8] M. Lewis, Y. Liu, N. Goyal, et al., "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in Proc. ACL, 2020.
[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, pp. 785–794, 2016.
[10] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[11] Tweepy Developers, "Tweepy: Twitter for Python," [Online]. Available: https://www.tweepy.org
[12] Hugging Face, "Transformers: State-of-the-art Natural Language Processing for PyTorch and TensorFlow 2.0," [Online]. Available: https://huggingface.co/transformers




10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY