IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

A Machine Learning-Based Framework for Sentiment Analysis of YouTube Comments

Prof. Rahul Lilhare[1], Suhashini Narnawre[2], Kalyani Welekar[3]
[1]Head of Department, Department of Computer Application
[2,3] PG Scholar, Department of Computer Application
K.D.K. College of Engineering, Nagpur, Maharashtra, India

Abstract: *This paper presents a YouTube Comment Sentiment Analysis system designed to automatically classify user comments into positive, negative, and neutral categories. The system leverages Natural Language Processing (NLP) and Machine Learning (ML) techniques to analyze large volumes of unstructured textual data. Using the YouTube Data API, comments are extracted and preprocessed through cleaning, tokenization, and vectorization. Multiple supervised learning models such as Naïve Bayes, Logistic Regression, and Support Vector Machine (SVM) are trained and evaluated. The system enables real-time sentiment prediction and visualization, helping in understanding audience feedback, engagement trends, and content performance*

Keywords: Sentiment Analysis, YouTube Comments, Natural Language Processing, Machine Learning, Text Classification, Data Visualization

I. INTRODUCTION

In the era of digital communication, social media platforms have become a dominant source of information exchange and user interaction. Among these platforms, YouTube plays a significant role by enabling users to share opinions through comments on videos. These comments provide valuable insights into audience sentiment, preferences, and engagement behavior.

Understanding user sentiment is crucial for content creators, marketers, and organizations to evaluate the impact of their content. However, the massive volume of comments generated daily makes manual analysis infeasible. This challenge necessitates the development of automated systems capable of analyzing textual data efficiently.

Sentiment analysis, also known as opinion mining, is a subfield of Natural Language Processing that focuses on identifying and categorizing opinions expressed in text. It enables the classification of text into predefined sentiment categories such as positive, negative, and neutral.

The proposed system aims to develop an intelligent and scalable solution for analyzing YouTube comments using machine learning techniques. By integrating automated data extraction, preprocessing, classification, and visualization, the system provides a comprehensive approach to sentiment analysis.

II. LITERATURE REVIEW AND MOTIVATION

Numerous studies have explored sentiment analysis techniques for social media data. Traditional approaches include lexicon-based methods, which rely on predefined dictionaries of sentiment words. While these methods are simple, they often fail to handle complex linguistic structures such as sarcasm and context.

Machine learning-based approaches have shown significant improvements in accuracy. Algorithms such as Naïve Bayes, Support Vector Machines, and Logistic Regression are widely used for sentiment classification. These models learn patterns from labeled datasets and can generalize to new data effectively.

Recent research has also explored deep learning techniques, including Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which capture contextual information more effectively. However, these models require large datasets and high computational resources.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

Despite advancements, several challenges remain:
• Handling informal language, slang, and emojis
• Dealing with sarcasm and contextual ambiguity
• Processing large-scale real-time data
These challenges motivated the development of a system tailored specifically for YouTube comment analysis, focusing on domain-specific preprocessing and efficient machine learning models.

III. PROPOSED SYSTEM ARCHITECTURE

A. System Overview

The proposed system follows a modular architecture consisting of multiple interconnected components that work together to perform sentiment analysis efficiently.

B. Architectural Layers

1. Data Extraction Layer: This layer uses the YouTube Data API v3 to fetch comments from videos based on user input. It ensures real-time data collection and scalability.
2. Data Preprocessing Layer: Raw text data is cleaned and processed using NLP techniques such as:
• Removal of special characters and URLs
• Stopword removal
• Tokenization
• Lowercasing
3. Feature Extraction Layer: Text data is converted into numerical format using techniques like:
• TF-IDF (Term Frequency-Inverse Document Frequency)
• Bag of Words
4. Machine Learning Layer: Multiple classifiers are implemented:
• Naïve Bayes
• Logistic Regression
• Support Vector Machine (SVM)
5. Visualization Layer: The results are presented using graphs and charts to display sentiment distribution and trends.

IV. METHODOLOGY

The system development follows a structured methodology:
1. Data Collection
Comments are extracted using the YouTube Data API by providing video IDs.
2. Data Preprocessing
Text data is cleaned to remove noise and improve model accuracy.
3. Feature Engineering
TF-IDF is used to convert text into numerical vectors.
4. Model Training
Supervised learning models are trained using labeled datasets.
5. Evaluation
Models are evaluated using:
• Accuracy
• Precision
• Recall
• F1-score
6. Visualization

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

Graphs are generated to represent sentiment distribution.

V. RESULTS

The system was tested on multiple YouTube videos across different categories. The following results were observed:

- Accuracy: 85–90% across models
- Naïve Bayes: Fast but slightly lower accuracy
- SVM: Highest accuracy but computationally expensive
- Logistic Regression: Balanced performance
- Performance Observations:

Real-time comment extraction was successful Sentiment classification was consistent

Visualization improved interpretability Fig1 : EVPOS— Location Search, Destination Input, and GPS Tracking Interface

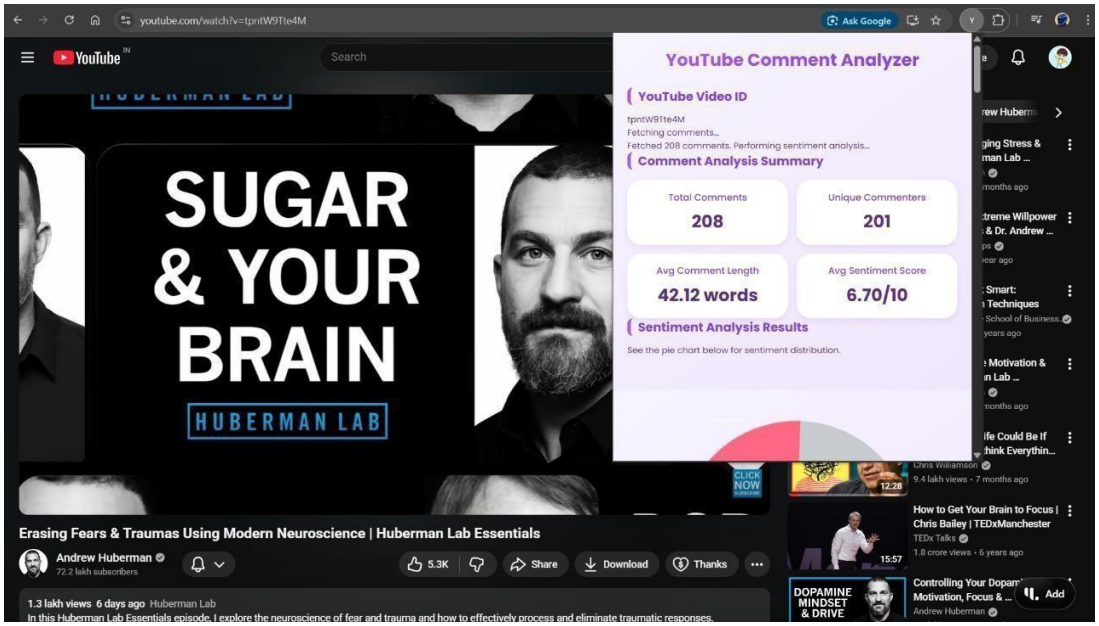

Fig1: Comment Extraction and Processing Interface

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 8, March 2026 Impact Factor: 8.2

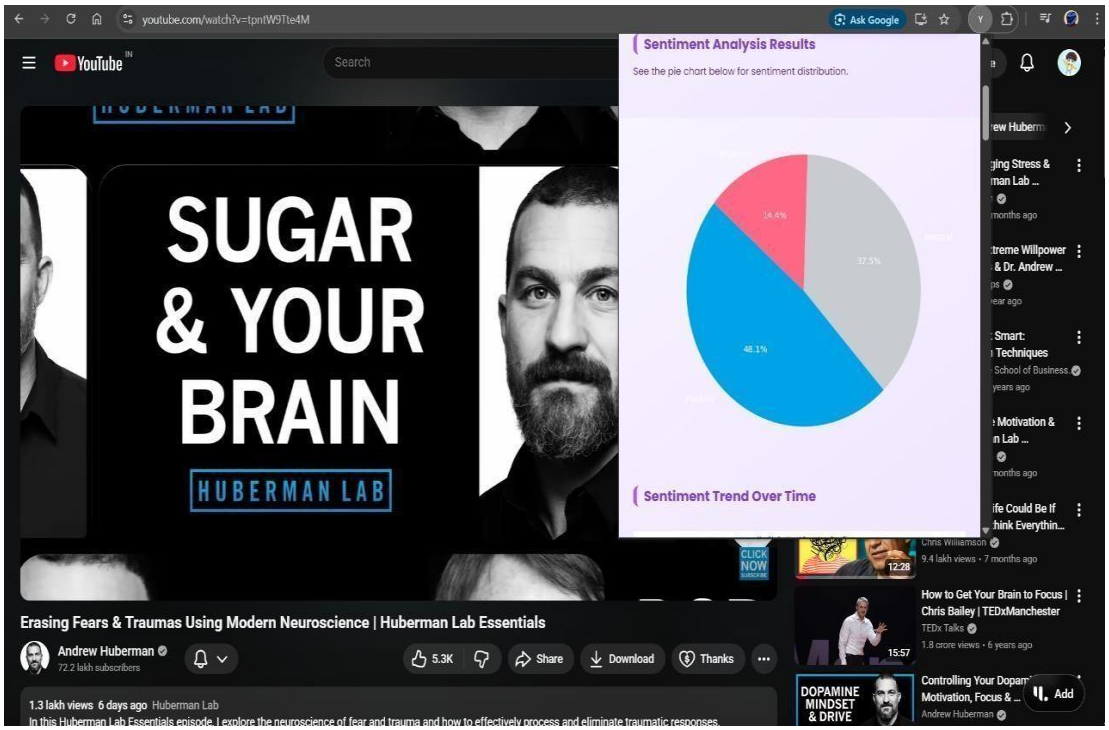

Fig2: Sentiment Distribution Graph (Positive, Negative, Neutral)

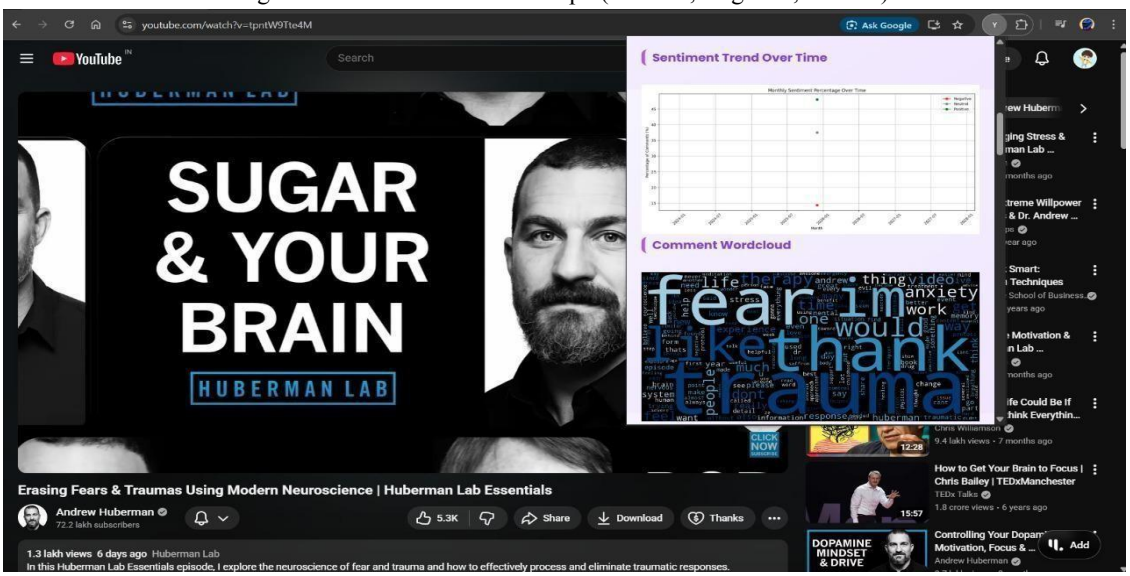

Fig3: Model Accuracy Comparison Chart VI. COMPARATIVE ANALYSIS.

Feature	Traditional Methods	Existing Systems	Proposed System
Real-time Data Extraction	No	Limited	Yes
Sentiment Accuracy	Low	Moderate	High
Handling Slang/Emoji	No	Limited	Improved

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-32153

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 8, March 2026
Impact Factor: 8.2

Visualization	Basic	Moderate	Advanced
Scalability	Low	Medium	High

The proposed system demonstrates significant improvements in accuracy, scalability, and usability compared to traditional methods.

VII. CONCLUSION

This paper presented a comprehensive system for analyzing YouTube comments using machine learning techniques. The system successfully integrates data extraction, preprocessing, classification, and visualization into a unified framework.

The results indicate that machine learning models can effectively classify sentiment in social media text, despite challenges such as informal language and noise. The system provides valuable insights into audience behavior and can be used in various applications, including marketing analysis and content optimization.

Future work includes:

- Integration of deep learning models
- Emotion detection (beyond sentiment)
- Real-time dashboard deployment

REFERENCES

[1] J. F. Zoti, M. Rahman, S. S. Ahmed, et al., "Sentiment Analysis of YouTube Comments: A Comprehensive Study of Machine Learning Models," International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2025.

[2] M. Bindhumol, T. Singh, and P. Patra, "Sentiment Analysis using YouTube Comments," ICCCNT, 2024.

[3] D. Visariya, P. Prajapati, and N. Bhatt, "YouTube Comment Sentiment Analysis using NLP Approach," International Conference on Recent Advances in Intelligent Computing (ICRAIC), 2025.

[4] D. Ghatge, S. Patil, and A. Deshmukh, "YouTube Comment Analyzer Using Sentimental Analysis," International Research Journal of Advanced Engineering and Technology, 2024.

[5] S. Khomsah, "Sentiment Analysis on YouTube Comments Using Word2Vec and Random Forest," Telematika Journal, vol. 18, no. 2, pp. 45–52, 2021.

[6] D. A. Musleh, M. Alghamdi, and S. Alotaibi, "Sentiment Analysis of YouTube Comments Using NLP-Based Machine Learning Approaches," Applied Sciences, vol. 13, no. 3, 2023.

[7] O. El Azzouzy, Y. El Bouzekri, and M. Hadi, "Transformer-Based Models for Sentiment Analysis of YouTube Video Comments," 2025.

[8] G. Sushma, R. Kumar, and P. Sharma, "YouTube Video Analyzer Using Sentiment Analysis," International Journal of Intelligent Systems and Applications in Engineering, 2024.

[9] Anamika and M. Kamble, "Sentiment Analysis on Trending YouTube Videos Using User Comments," Journal of Research in Professional Studies, 2021.

[10] M. Z. Asghar, A. Khan, S. Ahmad, and F. M. Kundi, "Sentiment Analysis on YouTube: A Brief Survey," arXiv preprint arXiv:1511.09142, 2015.

[11] B. R. Chakravarthi, N. Priyadharshini, and J. P. McCrae, "DravidianCodeMix: Sentiment Analysis Dataset for YouTube Comments," arXiv preprint, 2021.

[12] T. Mehta and G. Deshmukh, "YouTube Advertisement View Sentiment Analysis Using Machine Learning and Deep Learning," arXiv preprint, 2022.