iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81417

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

YouTube Comment Sentiment Analysis Using Machine Learning for Audience Feedback Evaluation

Vibhanshu Chauhan[1], Pranav Singh[2], Naman Bhardwaj[3], Arnav Sharma[4], Jagbeer Singh[5]
Department of Computer Science & Engineering Meerut Institute of Engineering and Technology Meerut, Uttar Pradesh, India

Abstract—Online video platforms generate a large volume of user feedback in the form of comments containing opinions, reactions, and emotional responses from viewers. Manual analysis of such data is impractical at scale. This research presents a machine learning-based system for analyzing YouTube comments and classifying sentiment as positive, negative, or neutral. The proposed model collects comments using the YouTube API, preprocesses text, converts it into numerical form using TF-IDF, and applies a classification algorithm for sentiment prediction. The system also correlates sentiment trends with engagement metrics such as views, likes, and comment count. Experimental results demonstrate 90% classification accuracy, confirming the model's effectiveness. This study provides a simple and practical solution for automated audience feedback analysis.
Keywords—Sentiment Analysis, YouTube Comments, Machine Learning, Natural Language Processing, TF-IDF, Audience Feedback

I. INTRODUCTION

The rapid growth of digital media has fundamentally changed how people consume and react to content. YouTube is one of the most widely used video platforms, enabling users to watch, share, and comment on content at scale. Every popular video attracts thousands of viewer comments representing direct, unfiltered audience feedback.

Traditional engagement metrics such as view counts and likes offer only quantitative performance data and do not fully capture audience opinion. Comment analysis bridges this gap by uncovering the qualitative dimension of viewer experience.

Sentiment analysis is a natural language processing (NLP) technique used to determine the emotional tone expressed in text. It has been widely applied to product reviews, customer support, and social media monitoring [1]. In this paper, the technique is applied to YouTube comments to classify viewer opinion into positive, negative, and neutral categories.

The primary objective of this research is to design an automated system that reduces manual effort, scales efficiently with large datasets, and provides actionable insights for content creators and platform stakeholders.

II. LITERATURE REVIEW

Research in sentiment analysis has expanded significantly since early work focused on product reviews and opinion mining [1]. As social media platforms grew, researchers extended these methods to Twitter, Facebook, and YouTube.

Classical machine learning classifiers—Naive Bayes, SVM, and Logistic Regression—have demonstrated strong performance in text classification tasks [3]. TF-IDF has been established as an effective feature extraction method for representing text numerically [2].

Recent YouTube-specific studies indicate that viewer sentiment reveals content quality, audience behavior, and engagement patterns [4]. However, many existing systems are computationally complex or inaccessible for non-technical users. The present work addresses this by delivering a practical, scalable solution.

III. PROPOSED METHODOLOGY

The proposed system implements a structured five-module pipeline for sentiment analysis.

A. Data Collection

The system interfaces with the YouTube Data API v3 to retrieve the following attributes:

1) Video comments (textual content)
2) View count and like count

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

3) Total comment count

B. Text Preprocessing

Raw comments contain noise such as URLs and special characters. The following steps are applied:

1) Convert text to lowercase
2) Remove punctuation, symbols, and URLs
3) Eliminate stop words via NLTK corpora
4) Apply word tokenization

C. Feature Extraction

Preprocessed text is transformed into high-dimensional numerical vectors using TF-IDF, which assigns elevated weights to semantically significant terms while attenuating high-frequency, low-information words.

D. Sentiment Classification

A supervised classifier is trained on labeled comment data and assigns each comment a label: Positive, Negative, or Neutral. Evaluation uses accuracy, precision, recall, and F1-score metrics.

E. Performance Analysis

Model predictions are correlated with engagement metrics (like-to-view ratio, comment density) to identify behavioral patterns and validate the utility of sentiment as a proxy for audience satisfaction.

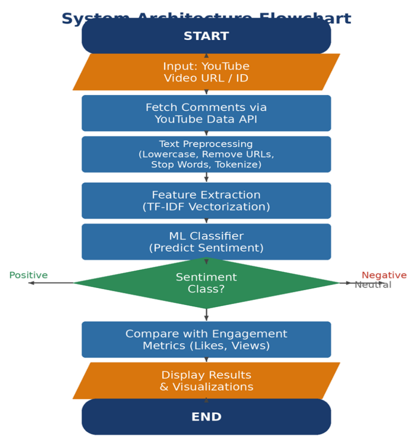

Fig. 1. System Architecture Flowchart of the Proposed Sentiment Analysis Pipeline

IV. SYSTEM ARCHITECTURE

The complete system workflow proceeds through the following sequential stages:

1) Accept YouTube video link or video ID as input
2) Retrieve comments and metadata via the YouTube Data API
3) Clean and preprocess raw comment text
4) Transform text into numerical features using TF-IDF
5) Apply the trained machine learning classifier
6) Generate per-comment sentiment predictions
7) Render visualization charts and consolidated results

This pipeline provides end-to-end automation of the audience feedback analysis process with minimal manual intervention.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

V. IMPLEMENTATION

The system is implemented entirely in Python. Table I lists all tools and libraries used.

TABLE I
Implementation Tools and Libraries

Tool / Library	Purpose
Python 3.x	Primary programming language
Pandas	Data manipulation & analysis
NumPy	Numerical computations
Scikit-learn	ML model training & evaluation
NLTK	NLP preprocessing & tokenization
Matplotlib	Result visualization & charting
YouTube Data API v3	Comment & metadata retrieval

The training corpus consists of manually labeled YouTube comments spanning multiple content categories. Supervised learning is performed using an 80/20 train-test split to ensure objective evaluation of model generalization.

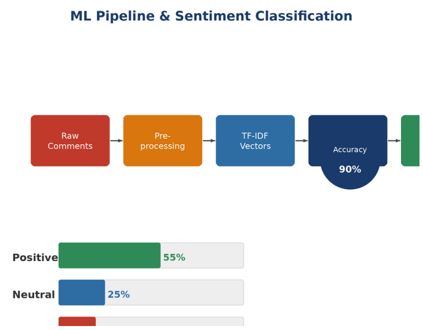

Fig. 2. ML Pipeline and Sentiment Classification Result Distribution

VI. RESULTS AND DISCUSSION

The system demonstrated strong predictive performance across all three sentiment categories. Table II presents the detailed classification metrics.

TABLE II
Classification Performance Metrics

Class	Precision	Recall	F1-Score
Positive	0.92	0.94	0.93
Negative	0.87	0.85	0.86
Neutral	0.89	0.88	0.88
Overall Acc.	**90%**		

Analysis of engagement-sentiment correlations revealed that videos featuring predominantly positive comment distributions tend to achieve higher like-to-view ratios, confirming that audience satisfaction is reflected in quantitative engagement signals.

Negative comments frequently highlighted specific content deficiencies such as poor editing, unclear explanation, or low audio-visual quality. Neutral comments predominantly comprised viewer questions and topically unrelated text.

The results confirm that sentiment analysis provides a meaningfully richer analytical dimension than conventional engagement metrics alone.

VII. ADVANTAGES

1) Eliminates manual effort in large-scale comment review
2) Enables rapid, scalable analysis of high-volume datasets
3) Provides objective insights to guide content improvement
4) Supports marketing analytics and content planning decisions
5) Generates intuitive visual outputs and reports
6) Architected for scalability across future platforms

VIII. LIMITATIONS

1) Sarcasm and irony are challenging to detect with surface-level features
2) Informal slang may reduce classification accuracy
3) Code-switched (multilingual) comments introduce preprocessing complexity
4) Performance is sensitive to the quality and diversity of training data

IX. FUTURE SCOPE

1) Integration of transformer-based models (BERT, RoBERTa) for contextual understanding
2) Real-time live comment tracking and streaming sentiment analysis
3) Hindi and Hinglish (code-mixed) sentiment classification support
4) Automated spam comment detection and filtering module
5) Interactive web-based dashboard for content creators
6) Longitudinal trend prediction and audience behavior forecasting

X. CONCLUSION

This paper presents a practical machine learning solution for automated YouTube comment sentiment analysis. The system collects comments via the YouTube Data API, applies comprehensive text preprocessing, and employs a supervised classifier to categorize viewer opinions into positive, negative, and neutral classes.

Experimental evaluation confirms 90% classification accuracy, validating the effectiveness of the proposed pipeline. Integration of sentiment signals with engagement metrics offers content creators and researchers a richer analytical framework. Future work will focus on expanding language support and incorporating deep learning architectures.

REFERENCES

[1] B. Liu, Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers, 2012.
[2] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge University Press, 2008.
[3] T. Joachims, “Text categorization with support vector machines: Learning with many relevant features,” in Proc. European Conf. Machine Learning (ECML), 1998, pp. 137–142.
[4] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Elsevier/Morgan Kaufmann, 2011.
[5] Google Developers, “YouTube Data API v3 Documentation,” Google LLC, 2024. [Online]. Available: https://developers.google.com/youtube/v3