Sentiment Analysis on YouTube Comments Using YouTube Data API V3

Bhuvansh Hiwrekar[1], Saksham Gorade[2], Shripad Bhakre[3], Tejas Wath[4], Shritesh Dharmale[5],

Guide - Prof. Pravin Kaware [6]

1,2,3,4,5,6Department of Computer Science & Engineering,
HVPM's College of Engineering & Technology, Amravati, Maharashtra, India

--************************----------------------------

Abstract:

In the era of digital media, YouTube has emerged as one of the most influential platforms for sharing information, entertainment, and public opinion. With billions of daily interactions, the platform generates enormous volumes of user-generated textual data in the form of comments. Understanding audience sentiment from these comments holds significant value for content creators, marketers, brand managers, and researchers. This paper presents a comprehensive study on automated Sentiment Analysis of YouTube comments using the YouTube Data API v3 for data collection and Natural Language Processing (NLP) techniques for classification. The proposed system fetches comment data from target YouTube videos, preprocesses the raw text, and employs multiple machine learning and deep learning models — including VADER, TextBlob, Support Vector Machine (SVM), Naive Bayes, and transformer-based BERT — to classify sentiments into Positive, Negative, and Neutral categories. Experimental results demonstrate that BERT achieves the highest classification accuracy of 93.1%, outperforming all traditional baselines. The system further provides interactive visual dashboards and trend reports to support data-driven decision-making.

Keywords — **BERT, Deep Learning, Machine Learning, Natural Language Processing, Opinion Mining, Sentiment Analysis, Social Media Mining, YouTube Data API v3.**

--************************----------------------------

I. INTRODUCTION

The rapid growth of the internet and social media platforms has fundamentally transformed how individuals communicate, consume content, and express opinions. YouTube, the world's largest video-sharing platform, hosts over two billion logged-in users monthly with over 500 hours of video uploaded every minute. The comment sections of YouTube videos serve as a rich and largely untapped repository of public opinion, emotional expression, and audience feedback. [1], [2]

Sentiment analysis, also referred to as opinion mining, is the computational study of identifying, extracting, and quantifying subjective information from textual data. The field lies at the intersection of Natural Language Processing (NLP), machine learning, and computational linguistics. Over the past decade, sentiment analysis has been applied across a wide range of domains including e-commerce product reviews, financial market prediction, healthcare feedback, and political opinion monitoring. [3]

Despite the enormous potential, manually reviewing thousands of YouTube comments is impractical, time-consuming, and prone to human bias. The YouTube Data API v3 provides a robust programmatic interface to fetch comment data at scale, making it an ideal data source for automated sentiment analysis. This paper proposes a complete Sentiment Analysis system integrating the YouTube Data API v3 with a multi-stage NLP pipeline and multiple classification models — lexicon-based (VADER, TextBlob), traditional ML (SVM, Naive Bayes), and transformer-based (BERT) — to classify sentiments as Positive, Negative, or Neutral. [4], [1]

II. Literature Review

A substantial body of research has been dedicated to sentiment analysis on social media platforms. This section reviews key works that directly informed the proposed system, organized by thematic focus.

Early investigations established that supervised classifiers including Naive Bayes, Maximum Entropy, and Support Vector Machines (SVM) could effectively predict document-level sentiment polarity in product reviews, laying the conceptual groundwork for all subsequent opinion mining research. [5]

Feature selection and weighting strategies were identified as critical determinants of classifier performance. Research by O'Keefe and Koprinska demonstrated that well-engineered linguistic features significantly improve model accuracy, underscoring the importance of preprocessing in the NLP pipeline. [2]

A comprehensive survey by Gunasekaran (2023) explored the full landscape of sentiment analysis techniques in NLP, covering lexicon-based, graph-based, machine learning, deep learning, ensemble-based, and hybrid methods, and identified key open challenges including sarcasm detection and multilingual sentiment analysis. [5]

Wankhade et al. (2022) conducted a detailed survey on sentiment analysis methods, applications, and challenges, systematically reviewing research from 1996 to 2022. The survey confirmed that transformer-based methods represent the current state-of-the-art across social media platforms. [19]

Salha Al Osaimi and Khan Muhammad Badruddin proposed an approach to predict sentiments in informal Arabic language using NLP and artificial intelligence. Their study highlighted the critical role of emotion icons (emojis) in developing accurate classifiers for informal social media text — directly applicable to YouTube comment analysis where emoji usage is pervasive. [6]

Pragya Tripathi, Santosh Kr Vishwakarma, and Ajay Lala (2015) built and compared two classifiers for sentiment analysis of English tweets using the RapidMiner platform. Their comparative evaluation methodology strongly influenced the multi-model evaluation design adopted in the present study. [7]

Abbi Nizar Muhammad, Saiful Bukhori, and Priza Pandunata (2019) applied a hybrid Naive Bayes-Support Vector Machine (NBSVM) classifier to classify YouTube comments. Using a 70:30 train-test split, the algorithm achieved a precision of 91% and a recall of 83%, demonstrating the effectiveness of ensemble approaches for YouTube comment classification. [4]

Musleh et al. (2023) presented a data collection, preprocessing, and modeling framework for sentiment analysis on Arabic YouTube comments. Six ML models were benchmarked on 4,212 labeled comments, with Naive Bayes achieving the highest accuracy of 94.62% and MCC score of 91.46%. [8]

Pichad et al. (2023) assessed out-of-the-box machine learning classifiers for trending YouTube video sentiment analysis, reporting approximately 75% accuracy for NLP-based methods [9]. Giri, Sirsath, and Kanakia (2024) conducted a comprehensive study classifying YouTube comments using multiple ML and NLP techniques at IEEE I2CT 2024, demonstrating strong correlations between user sentiment trends and real-world events. [10]

The development of BERT (Bidirectional Encoder Representations from Transformers) by Devlin et al. (2019) at Google represented a landmark advance in NLP. Through pre-training on large corpora with a bidirectional self-attention mechanism, BERT captures full contextual information and achieves state-of-the-art performance across NLP benchmarks. [13]

Nahas, Swetha, and Nandakumar (2024) confirmed that transformer-based architectures consistently outperform traditional ML models on YouTube data [14]. A recent ScienceDirect study (2025) comparing BERT, GPT, RoBERTa, and T5 for YouTube comment sentiment analysis concluded that fine-tuned BERT achieves among the highest performance scores in terms of robustness and generalizability. [15]

Song Qin, Ronaldo Menezes, and Marius Silaghi (2010) created a YouTube recommender network exploiting social network features, establishing a systematic API-based data mining framework. Malik and Tian (2017) further presented a scalable framework for collecting and processing YouTube metadata at scale, demonstrating the feasibility of large-scale YouTube data collection. [1], [17]

III. Problem Definition

YouTube is among the most widely used video-sharing platforms globally, with millions of users posting comments daily. These comments contain valuable insights for content creators, marketers, and researchers. However, manually analyzing thousands of comments is time-consuming, inefficient, and inherently prone to human bias and inconsistency. [1], [2]

The key challenges in building an automated YouTube comment sentiment analysis system include:

- Efficiently collecting large comment volumes in real-time from YouTube's vast content ecosystem. [1]
- Processing noisy, unstructured text containing slang, emojis, abbreviations, code-switching, and mixed languages. [6], [4]
- Accurately classifying sentiments (Positive, Negative, Neutral) while handling linguistic ambiguity, negation, and context. [4], [5]
- Scaling the system across channels, topics, and categories without extensive retraining. [3], [18]
- Presenting results in intuitive visual formats for non-technical stakeholders. [7], [9]

Currently, no comprehensive automated system exists that integrates the YouTube Data API v3 with state-of-the-art NLP and deep learning to extract, analyze, and visualize YouTube comment sentiments in a unified platform. This project directly fills that gap. [4], [1], [10]

IV. Proposed Methodology

The proposed system performs end-to-end sentiment analysis by integrating the YouTube Data API v3 with a four-phase pipeline: (1) data collection, (2) data preprocessing, (3) sentiment classification, and (4) result visualization.

1. Data Collection

The YouTube Data API v3 programmatically fetches comments from specified videos, playlists, or channels, retrieving comment text, author username, publication timestamp, like count, and reply count. The API supports pagination via page tokens, enabling extraction of large comment volumes (up to 100 per API call). Collected data is stored in CSV format for analysis and MySQL for persistent queryable storage. API authentication uses OAuth 2.0 and API key mechanisms per Google's developer guidelines. [1], [17]

2. Data Preprocessing

Raw YouTube comment text is inherently noisy and requires a comprehensive preprocessing pipeline before model input. The following sequential steps are applied: [4], [3]

(1) URL and HTML tag removal using regular expressions. (2) Special character normalization. (3) Emoji handling: emojis are mapped to textual descriptors (e.g., 'heart' → 'love') using the emoji Python library, preserving their sentiment signal. [6]

(4) Case normalization to lowercase. (5) Stopword removal using NLTK's corpus. [3]

(6) Tokenization into individual tokens. (7) Stemming and lemmatization using the Porter Stemmer and WordNet Lemmatizer from NLTK to reduce vocabulary dimensionality. [7]

3. Sentiment Classification

Preprocessed comments are passed through a multi-model classification pipeline across three paradigms:

Lexicon-Based Models: VADER (Valence Aware Dictionary and sEntiment Reasoner) is a rule-based tool designed for social media text, outputting a compound polarity score from -1 to +1. TextBlob provides polarity and subjectivity scores. These lightweight models serve as baselines requiring no training data. [7], [19]

Traditional ML Classifiers: TF-IDF vectorization feeds Naive Bayes and SVM classifiers trained on labeled YouTube comment datasets. Hyperparameter tuning uses grid search with 5-fold cross-validation. [4], [3], [20]

Transformer-Based Deep Learning: BERT (bert-base-uncased) is fine-tuned using the Hugging Face Transformers library with a learning rate of 2e-5, batch size of 32, and 4 training epochs. BERT's bidirectional attention mechanism captures the full contextual meaning of each comment. [13], [14], [15]

4. Result Visualization

Classification outputs are presented through interactive visual dashboards built with Plotly, served via a Flask web backend. Visualizations include sentiment distribution pie charts, time-series trend graphs, per-sentiment word clouds, and per-video or per-channel comparative reports. [1], [18]

V. System Requirements

Software: Python 3.10+, Flask (backend), React.js (frontend), TensorFlow 2.x, scikit-learn, Hugging Face Transformers, NLTK, spaCy, VADER, TextBlob, YouTube Data API v3, MySQL 8.0, Plotly, Matplotlib, AWS EC2 / GCP.

Hardware: Minimum 16 GB RAM (32 GB recommended for BERT fine-tuning); NVIDIA CUDA-compatible GPU (minimum 8 GB VRAM); 50 GB SSD; stable broadband internet connection.

VI. Experimental Results and Discussion

A. Dataset

The experimental dataset was compiled by extracting comments from 50 YouTube videos spanning five content categories: Technology, Entertainment, Education, News, and Gaming. A total of 45,000 comments were collected using the YouTube Data API v3. Comments were manually labeled by three independent annotators via majority-vote into three classes: Positive (52%), Negative (28%), and Neutral (20%). Inter-annotator agreement was measured using Cohen's Kappa coefficient ($\kappa = 0.79$), indicating substantial agreement. An 80:20 train-test split was applied. [4], [1], [8]

B. Model Performance Comparison

Table I summarizes the classification accuracy, precision, recall, and F1-score achieved by each model on the held-out test set of 9,000 comments. [7], [4], [9]

Result Table

TABLE I: Performance Comparison of Sentiment Classification Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
VADER (Lexicon)	72.4	70.1	72.4	71.0
TextBlob (Lexicon)	68.9	67.2	68.9	67.8
Naive Bayes	79.3	78.5	79.3	78.7
SVM (TF-IDF)	84.7	83.9	84.7	84.2
BERT (Fine-tuned)	93.1	92.8	93.1	92.9

C. Discussion

The results clearly demonstrate a progressive improvement moving from lexicon-based methods to traditional ML to transformer-based deep learning. VADER and TextBlob achieve lower accuracy (72.4% and 68.9%) due to limited handling of contextual nuances, slang, and negation characteristic of YouTube comments. [7], [19]

The SVM classifier with TF-IDF features achieves 84.7% accuracy, confirming findings from multiple prior studies that traditional ML with well-engineered features remains competitive. This is consistent with the NBSVM approach of Muhammad et al. [4] which also reported strong SVM performance on YouTube data. [4], [20]

The fine-tuned BERT model achieves the highest accuracy of 93.1% with an F1-score of 92.9%. BERT's bidirectional attention mechanism enables it to capture full comment context, making it particularly effective at resolving linguistic ambiguity in short, informal text. [13], [14], [15]

Analysis of misclassified examples reveals that sarcasm and irony represent the primary failure mode across all five models — cases where surface-level positive language conveys a genuinely negative sentiment. This finding is consistent with the broader literature identifying sarcasm detection as one of the most persistent open challenges in social media sentiment analysis. [6], [5], [16]

VII. Desired Implications

The proposed system delivers tangible practical benefits across stakeholder groups:

1) Enhanced Content Strategy: Content creators gain quantified insight into audience feedback, enabling them to tailor future videos to better meet viewer expectations. [1], [18]

2) Efficient Sentiment Monitoring: The automated pipeline replaces manual comment review with a scalable, near-real-time system capable of processing thousands of comments per minute. [4], [9]

3) Business and Brand Intelligence: Organizations can continuously monitor brand perception and customer opinion when their products or services appear in YouTube content. [3], [8]

4) Data-Driven Decision Making: Interactive sentiment dashboards enable stakeholders at all levels to make evidence-based decisions grounded in real audience data. [7], [11]

5) Scalable and Reusable Architecture: The modular system design enables adaptation to different channels, domains, and languages, and can be extended to other social media platforms. [6], [10]

6) Academic Contributions: The comparative evaluation across all three model paradigms under identical experimental conditions provides a reproducible benchmark for future NLP research. [19], [5]

VIII-Conclusion

This paper presented a comprehensive, end-to-end automated Sentiment Analysis system for YouTube comments, integrating the YouTube Data API v3 with a multi-stage NLP preprocessing pipeline and a rigorous comparative evaluation of five classification models. Experimental results on 45,000 YouTube comments across five content categories demonstrate that the fine-tuned BERT transformer model achieves the highest overall performance (93.1% accuracy, 92.9% F1-score), significantly outperforming lexicon-based baselines and traditional ML classifiers. The system's interactive visual dashboards make sentiment insights accessible to non-technical users, enabling practical deployment across content creation, brand monitoring, and research applications. [6], [7], [4], [1], [3], [8], [13]

Future work directions include: (1) incorporating dedicated sarcasm and irony detection modules via multi-task learning, (2) extending to multilingual analysis using mBERT, (3) integrating aspect-based sentiment analysis for finer-grained feedback extraction, (4) real-time streaming comment analysis using Apache Kafka for live video events, and (5) incorporating multimodal signals such as reply thread context and video metadata to further improve classification accuracy. [5], [14], [16], [18]

REFERENCES

[1] S. Qin, R. Menezes, and M. Silaghi, "A Recommender System for YouTube Based on its Network of Reviewers," in Proc. IEEE Second International Conference on Social Computing, 2010.

[2] T. O'Keefe and I. Koprinska, "Feature Selection and Weighting in Sentiment Analysis," in Proc. 14th Australasian Document Computing Symposium, Sydney, Australia, December 2009.

[3] W. Yang and Z. Qian, "Understanding the Characteristics of Category-Specific YouTube Videos," December 2011, Available: CiteSeerX.

[4] A. N. Muhammad, S. Bukhori, and P. Pandunata, "Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes–Support Vector Machine (NBSVM) Classifier," in Proc. International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 2019.

[5] K. P. Gunasekaran, "Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review," arXiv preprint arXiv:2305.14842, May 2023.

[6] S. Al Osaimi and K. M. Badruddin, "Sentiment Analysis of Arabic Tweets Using RapidMiner," Dept. of Information System, Imam Muhammad ibn Saud Islamic University, KSA.

[7] P. Tripathi, S. K. Vishwakarma, and A. Lala, "Sentiment Analysis of English Tweet Using Rapidminer," in Proc. International Conference on Computational Intelligence and Communication Networks, 2015, pp. 668–672.

[8] D. A. Musleh, I. Alkhwaja, A. Alkhwaja, M. Alghamdi, H. Abahussain, F. Alfawaz, N. Min-Allah, and M. M. Abdulqader, "Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation," Big Data and Cognitive Computing, MDPI, vol. 7, no. 3, p. 127, July 2023. DOI: 10.3390/bdcc7030127.

[9] S. Pichad, S. Kamble, R. Kalamb, and S. Chavan, "Analysing Sentiments for YouTube Comments Using Machine Learning," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 11, issue V, May 2023.

[10] R. Giri, M. Sirsath, and H. T. Kanakia, "Youtube Comments Sentiment Analysis," in Proc. IEEE 9th International Conference for Convergence in Technology (I2CT), 2024, pp. 1–4.

[11] A. S. Aiswarya and H. Rajeev, "Youtube Comment Sentimental Analysis," Indian Journal of Data Mining (IJDM), vol. 4, issue 1, May 2024. DOI: 10.2139/ssrn.4846213.

[12] J. L. Rachel, M. Sawood, A. Khan, and B. P. Kumar, "Extracting Sentiments from YouTube Comments Using Deep Learning and Transformers," International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2024. DOI: 10.22214/ijraset.2023.57783.

[13] 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, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

[14] N. Nahas, P. Swetha, and R. Nandakumar, "Classifiers for Sentiment Analysis of YouTube Comments: A Comparative Study," in Proc. International Conference on Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication (MDCWC 2023), Springer, 2024. DOI: 10.1007/978-3-031-47942-7_38.

[15] "Transformer-based models for sentiment analysis of YouTube video comments," ScienceDirect, 2025. DOI: 10.1016/j.array.2025.x.

[16] A. Shevtsov et al., "TWSSenti: A Novel Hybrid Framework for Topic-Wise Sentiment Analysis on Social Media Using Transformer Models," arXiv preprint arXiv:2504.09896, April 2025.

[17] H. Malik and Z. Tian, "A Framework for Collecting YouTube Meta-Data," Procedia Computer Science, vol. 113, pp. 194–201, 2017. DOI: 10.1016/j.procs.2017.08.347.

[18] P. Surana, S. Vispute, M. Kalbhor, P. Ugale, I. Syeda, M. Yakumsha, and A. Suryawanshi, "Opinion Mining of YouTube Video Comments Using Machine Learning and Deep Learning," Springer, 2026. DOI: 10.1007/978-3-032-06694-7_13.

[19] K. Wankhade, A. C. S. Rao, and C. Kulkarni, "A Survey on Sentiment Analysis Methods, Applications, and Challenges," Artificial Intelligence Review, Springer, vol. 55, no. 7, pp. 5731–5780, 2022.

[20] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," in Proc. EMNLP 2002, pp. 79–86.