



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 8, August 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Fake YouTube Review Detection using Abstract based Sentiment Analysis

**Prof. Shraddha Ovale, Prajwal Kadam, Ankita Sawant, Vinay Chopade, Priyanka Kaldate**

Professor, Department of Computer Engineering, Pimpri Chinchwad College of Engineering Pune, India

U.G. Student, Department of Computer Engineering, Pimpri Chinchwad College of Engineering Pune, India

U.G. Student, Department of Computer Engineering, Pimpri Chinchwad College of Engineering Pune, India

U.G. Student, Department of Computer Engineering, Pimpri Chinchwad College of Engineering Pune, India

U.G. Student, Department of Computer Engineering, Pimpri Chinchwad College of Engineering Pune, India

**ABSTRACT:** Fake video identification on social media platforms is a genuine issue today, and it has become increasingly widespread and difficult to identify. The detection of it in its early stages is a serious difficulty. Another issue in false video identification is the scarcity of tagged data for the model. To address the previously mentioned problem, this work provides a novel false video detection framework based on aspect-based sentiment analysis. The information from the video reviews from the social environment is used by the purpose framework. In the first phase, sentiment analysis will be used on user reviews to determine if they are positive, negative, or neutral. Second, NLP will be used to extract text features from the imbalance review dataset. Finally, model fine tuning will be performed using ATEPC.

**KEYWORDS:** YouTube, App Script, API, Machine Learning, Data mining, Sentiment Analysis, Aspect Based Sentiment Analysis, Lemanization, WordNetLemmatizer, WordCloud, Natural Language Processing, Python Aspect-Based Sentiment Analysis, Aspect Term Extraction and Sentiment Classification.

## I. INTRODUCTION

Indeed, as online technologies continue to advance, more and more platforms—like blogs, social networks, forums, and website reviews—are opening up to accept user-generated material. For those engaged in data mining, both people and organizations, this flood of data offers possibilities and problems. Analysis and understanding of the large amounts of unstructured text data in natural languages are severely hampered. Gaining insight into human behavior, beliefs, and ideas is an innate urge, and the emergence of online applications such as blogs, social networks, e-commerce sites, and other communication platforms has made this possible in ways never before possible. However, automated methods for organization and analysis are required due to the massive amount of data created by these platforms.

Sentiment analysis is a crucial task in Natural Language Processing (NLP) for determining general sentiment towards entities, consumer feedback, and public opinion. Professionals have given it a lot of attention. It involves identifying sentiment in textual data using computer approaches, allowing for automatic interpretation of dominant attitudes and emotions. Within NLP, Aspect Based Sentiment Analysis (ABSA) offers a more detailed analysis by detecting qualities or features in text and ascertaining associated sentiment, addressing the limitations of standard sentiment analysis which assigns an overall emotion to text.

ABSA involves extracting opinions and points from text to understand beliefs and viewpoints. Opinion Target Extraction (OTE) identifies words or phrases related to qualities being discussed, known as targets. Aspect Category Detection (ACD) classifies aspects mentioned in text, such as pricing, ambiance, meal quality, and service in a restaurant review, assigning objectives to aspect groups for analysis. This method allows for a deeper understanding of opinions expressed in the text and provides a more complex comprehension of topics discussed, like the cost or quality of a meal in a restaurant review.

## II. RELATED WORK

Singh and Tiwari [1] explore sentiment analysis within YouTube comments, unveiling intricate user sentiment dynamics and their correlation with real-world events. Athar [2] extends sentiment analysis to scientific citations, employing Naive Bayes and Support Vector Machine classifiers to enrich scholarly discourse. Kumar et al. [3] contribute to ABSA methodologies, emphasizing AI-driven approaches over manual feature extraction, while Balli et al. [4] underscore the importance of supervised learning in sentiment analysis across linguistic contexts, focusing on Turkish Twitter content. Kalbhor et al. [6] advance sentiment analysis through hybrid models, achieving notable accuracy on SemEval-2014 Task 4. Kaur et al. [7] showcase exceptional results across various datasets using deep learning and Machine Learning techniques, highlighting challenges in negation handling and dataset diversity. Tayal et al. [8] make significant progress in ensemble learning for haptic feedback, enhancing model performance. Zhou Gui Zhou [9] explores feature selection methodologies, achieving respectable F1 scores but identifying areas for improvement.

Yang et al. [11] and Lv et al. [12] tackle sentiment analysis on news and diverse datasets, respectively, with varying degrees of success and challenges. Transitioning to 2021, Yan et al. [13] achieve impressive accuracy in sentiment analysis using deep learning on the Weibo senti 100k dataset, while Rietzler et al. [14] and Li et al. [15] delve into aspect-target sentiment classification, pushing the boundaries of ABSA. Aydin and Gungor [16] refine ABSA frameworks for capturing neutral sentiments and adapting to constituency parser-based models, reflecting a commitment to methodological sophistication and improvement

## III. METHODOLOGY

### A. Proposed System Methodology

#### 1. Data Creation

The Proposed work uses YouTube API v3.0 to gather comments from Bollywood film genre videos, focusing on one video to extract data efficiently. Researchers limit comments to 100 per video for research parameters. Insights are gained from analyzing top-rated comments for viewer opinions and engagement trends in Bollywood movies.

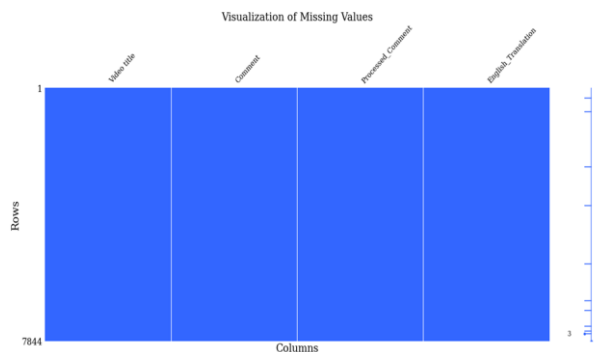


Fig 1. Identifying Missing Entries

#### 2. Data Preprocessing

Data preprocessing for YouTube comments involves standardizing and cleaning text by changing to lowercase, standardizing Unicode, removing links, special characters, punctuation, numbers, duplicates, line breaks, excessive spaces, common words, and lemmatization for consistency, legibility, and insights extraction.

**A) Lowercasing:** Convert all text to lowercase letters to ensure consistency and accurate analysis regardless of capitalization.

**B) Unicode Normalization:** Unicode normalization is important for text with characters from multiple languages and sets, ensuring consistency and compatibility.

**C) Hyperlink Removal:** YouTube comments often include irrelevant hyperlinks, so we remove them to focus on text analysis.



- D) **Special Character Removal:** Emojis, symbols, and non-standard punctuation can distract from text meaning, cleaned for clearer analysis.
- E) **Repeated Character Reduction:** Repeated characters used online for emphasis can distort text; reducing them ensures analysis consistency.
- F) **Whitespace Standardization:** Inconsistent whitespace in text is standardized by removing extra spaces for clarity and consistency.
- G) **Punctuation Removal:** Punctuation aids written communication but isn't crucial for online content analysis; removing can improve understanding.
- H) **Numeric Digit Removal:** Removing numbers allows for a focus solely on linguistic content without numerical distractions.
- I) **Special Character Replacement:** special characters and symbols can complicate processing or encoding in text analysis algorithms.
- J) **Lemmatization:** Lemmatization reduces words to base form for accurate analysis and interpretation of text.
- K) **Stopword Removal:** Removing common words like "and," "the," or "is" improves analysis quality and relevance.

Comment	Processed_Comment
Wow! Bahut kuch kaha Angrezon ka Sahara le kar ðŸ™,  Samajhne waale samajh gaye ...ðŸ™ ...	wow bahut kuch kaha angrezon ka sahara le kar brsamajhne waale samajh gaye

Fig 2. Data Preprocessing

### 3. Translating Text (Hinglish to English)

The plan to spot fake reviews on YouTube relies heavily on converting Hinglish comments to English to detect inaccuracies. Many users write comments in Hinglish, a mix of Hindi and English, which can create misunderstandings and complicate emotion and idea comprehension. To tackle this issue, we created the vital `hinglish\_to\_english` function. This function is crucial in translating Hinglish comments accurately, ensuring the reliability of our sentiment analysis and aspect extraction systems. Our rigorous verification process checks for errors in the input text to maintain information integrity during translation. Using services like Google Translate, our system seamlessly converts Hinglish to English, uncovering valuable insights. In the event of errors, our system captures and records error messages for analysis and improvement. By proactively addressing issues, we guarantee the effectiveness and dependability of our translation process, enabling us to handle potential challenges effortlessly.

Processed_Comment	English_Translation
copy kyu kartha vhai app na kudha ka banaa bollywood hamesha dusre ka copy paste kartha hai	Why does the Bollywood app always paste the copy of the other.

Fig 3. Result of Translation



7. Polarity

The calculate\_polarity function is essential in our sentiment analysis system for evaluating written text sentiment, especially in detecting fake YouTube reviews. The system first verifies the input text format, checking for correct data types like floating-point numbers. It assigns a neutral sentiment score of 0 to inputs containing decimal errors to maintain reliability. Valid inputs undergo sentiment polarity calculation using a sentiment analyzer tool, extracting a compound sentiment score for overall sentiment interpretation. This function assesses the text's emotional tone by determining positive, negative, or neutral feelings, aiding in identifying misleading content and enhancing sentiment analysis accuracy.

English_Translation	Polarity
Akshay Kumar Love Comedy	0.7717
The angle of comment has come like me	0.3612
Look at the angle	0
angle march acting	0
love people nice	0.7906
Wrong Movie	-0.4767

Fig 6. Polarity

8. Aspect Terms Sentiment

The extract\_aspects\_and\_sentiments function is pivotal in the proposed framework for analyzing sentiments in YouTube comments, providing valuable user input to understand viewpoints better. Using Aspect-Based Sentiment Analysis (ABSA), the function uncovers key aspects and associated sentiments, facilitating specific insights gathering from text data. Initially, an aspect\_sentiment\_results list is created to store extraction outcomes. The ABSA model, trained extensively on English text, identifies aspects and sentiments accurately. The function evaluates comments from YouTube reviews meticulously, using a progress bar for feedback. Detected aspects and sentiments are compiled into a dictionary in aspect\_sentiment\_results. This information is integrated into a DataFrame to enhance understanding for improving fake review detection systems efficiently.

Video title	Comment	Processed Comment	English Translation	Tokenized POS	Polarity	Aspect_Sentiment
Tu Jhoothi Main Makkaar Full Movie Hindi   Ranbir Kapoor, Shraddha Kapoor	Nice movie	nice movie	nice movie	[('nice', 'ADJ'), ('movie', 'NOUN')]	0.4215	{'Aspect': ['movie'], 'Sentiment': ['Positive']}

Fig 7. Result of Aspect-Sentiment Extraction.

B. Proposed System Diagram

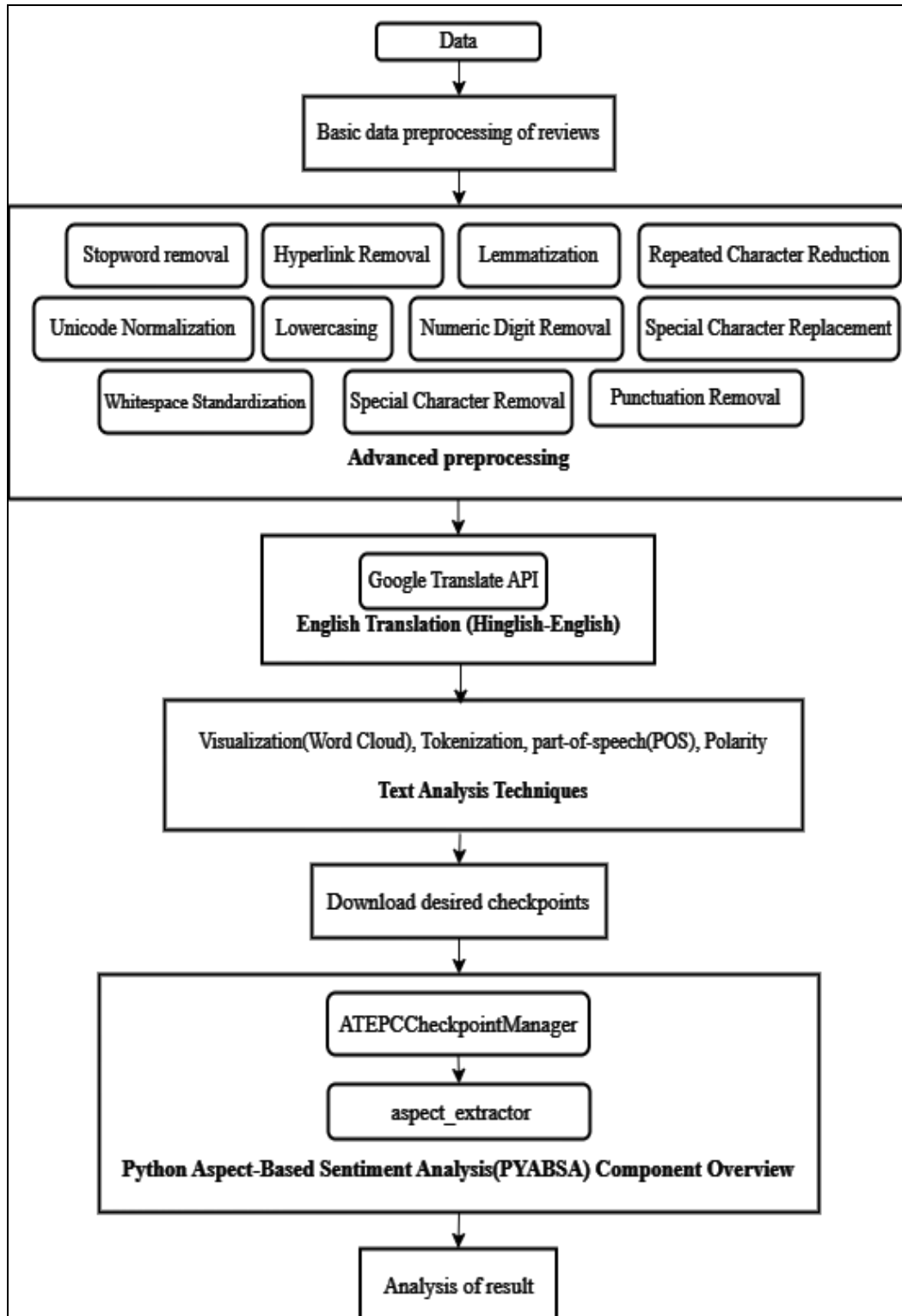


Fig 8. Proposed System

#### IV. EXPERIMENTAL RESULTS

In a study on detecting fake YouTube reviews, advanced natural language processing methods were used to analyze user feedback. The PyABSA library, with its pre-trained models, played a key role in extracting aspects and classifying sentiments. The ATEPC aspect extractor identified main topics and sentiments in comments, aiding in the classification of emotions as positive, negative, or neutral. This streamlined process improved the identification of fake reviews by focusing on frequently mentioned aspects and associated sentiments, enhancing the credibility of online review platforms.

#### V. CONCLUSION

Utilizing advanced natural language processing techniques like aspect extraction and sentiment analysis, our research successfully detects and reduces fake reviews in YouTube comments. PyABSA's ATEPC aspect extractor reveals key topics and user emotions, highlighting the importance of strong sentiment analysis to detect misleading online activities. Ongoing efforts are needed to combat data biases and improve detection systems for trustworthy online reviews, promoting trust and transparency in digital platforms.

#### REFERENCES

1. Ritika Singh, Ayushka Tiwari, "YouTube Comments Sentiment Analysis", International Journal of Scientific Research in Engineering and Management (IJSREM), Volume 05, Issue 05, May - 2021. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
2. Awais Athar, "Sentiment Analysis of Scientific Citations", University of Cambridge, Computer Laboratory, Tech Report No. UCAM-CL-TR-856, June 2014. DOI: 10.48456/tr-856. URL: <https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-856.pdf>.
3. Dinesh Kumar, Avdhesh Gupta, Vishan Kumar Gupta, Ashotosh Gupta, "Aspect-Based Sentiment Analysis using Machine Learning and Deep Learning Approaches", International Journal on Recent and Innovation Trends in Computing and Communication, Volume 11, Issue 5s, ISSN: 2321-8169 DOI: 10.17762/ijritcc.v11i5s.6636.
4. Cagla Balli, Mehmet Serdar Guzel, Erkan Bostanci, and Alok Mishra, "Sentimental Analysis of Twitter Users from Turkish Content with Natural Language Processing", Computational Intelligence and Neuroscience, Volume 2022, Article ID 2455160, 17 pages, DOI: 10.1155/2022/2455160.
5. Azadeh Mohammadi, Anis Shaverizade, "Ensemble Deep Learning for Aspect-based Sentiment Analysis", International Journal of Nonlinear Analysis and Applications (Int. J. Nonlinear Anal. Appl.), Volume 12, Special Issue, Winter and Spring 2021, Pages 29-38, ISSN: 2008-6822 (electronic), DOI: 10.22075/IJNAA.2021.4769.
6. Shraddha Kalbhor, Dinesh Goyal, Kriti Sankhla ; Taming Misinformation: Fake Review Detection on Social Media platform using Hybrid Ensemble Technique. International Journal of Electrical and Electronics Research (IJEER) , Volume 10, Issue 2
7. J Kaur, G., Sharma, A. A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. J Big Data 10, 5 (2023). <https://doi.org/10.1186/s40537-022-00680-6>
8. Tayal, Madhuri A. et al. "VMLHST: Development of an Efficient Novel Virtual Reality ML Framework with Haptic Feedbacks for Improving Sports Training Scenarios." International Journal of Electrical and Electronics Research (2023): n. pag.
9. Zhou Gui Zhou " Research on Sentiment Analysis Model of Short Text Based on Deep Learning" Hindawi Scientific Programming Volume 2022, Article ID 2681533, 7 pages <https://doi.org/10.1155/2022/2681533>
10. M. Yang, B. Jiang, Y. Wang, T. Hao , Y. Liu " News Text Mining-Based Business Sentiment Analysis and Its Significance in Economy" Frontiers in Psychology Volume 2022 <https://doi.org/10.3389/fpsyg.2022.918447>
11. Lv, Yanxia& Wei, Fangna& Cao, Lihong& Peng, Sancheng Niu, Jianwei& Yu, Shui& Wang, Cuirong. (2020). Aspect-level Sentiment Analysis Using Context and Aspect Memory Network. Neurocomputing. 428. 10.1016/j.neucom.2020.11.049.
12. Yan, Wei & Zhou, Lifan& Qian, Zhengjiang& Xiao, Le & Zhu, Haixia. (2021). Sentiment Analysis of Student Texts Using the CNN-BiGRUAT Model. Scientific Programming. 2021. 10.1155/2021/8405623.
13. A. Rietzler, S. Stabinger, P. Opitz and S. Engl (2020) " Adapt or Get Left Behind: Domain Adaptation through BERT Language Model Finetuning for Aspect-Target Sentiment Classification " <https://doi.org/10.48550/arXiv.1908.11860>



14. J Li, X., Fu, X., Xu, G., Yang, Y., Wang, J., Jin, L., ... Xiang, T. (2020). Enhancing BERT Representation With Context-aware Embedding For Aspect-Based Sentiment Analysis. IEEE Access, 1–1. doi:10.1109/access.2020.297851110.1109/access.2020.2978511 downloaded on 2020-04-26
15. Aydin, C. R., & Gungor, T. (2020). Combination of Recursive and Recurrent Neural Networks for Aspect-Based Sentiment Analysis Using Inter-Aspect Relations. IEEE Access, 8, 77820–77832. doi:10.1109/access.2020.2990306
16. J. Huang, H. Rui, G. Li, X. Qu, T. Tao and X. Zheng, "Multi-Label Learning With Hidden Labels," in IEEE Access, vol. 8, pp. 29667- 29676, 2020, doi: 10.1109/ACCESS.2020.29725



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details