



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 12, Issue 4, April 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

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

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



# Sentiment Analysis in Youtube Videos using Machine Learning

Sarthak Zende<sup>1</sup>, Atharwa Wadkar<sup>2</sup>, Aayush Giri<sup>3</sup>, Mujahid Hussain<sup>4</sup>, Suresh Shinde<sup>5</sup>

U.G. Student, Department of Computer Engineering, SKN College of Engineering, Pune, Maharashtra, India<sup>1</sup>

U.G. Student, Department of Computer Engineering, SKN College of Engineering, Pune, Maharashtra, India<sup>2</sup>

U.G. Student, Department of Computer Engineering, SKN College of Engineering, Pune, Maharashtra, India<sup>3</sup>

U.G. Student, Department of Computer Engineering, SKN College of Engineering, Pune, Maharashtra, India<sup>4</sup>

Associate Professor, Department of Computer Engineering, SKN College of Engineering, Pune, Maharashtra, India<sup>5</sup>

**ABSTRACT:** This research paper investigates sentiment analysis in YouTube videos using machine learning. With the explosive growth of online video content, it has become increasingly important to understand how viewers respond to videos. Sentiment analysis is a powerful tool that can help us understand the emotions and attitudes expressed in videos. In this paper, we explore the effectiveness of machine learning algorithms in analyzing the sentiment of YouTube videos. Our purpose is to provide a comprehensive analysis of the current state of sentiment analysis in YouTube videos and to identify areas for future research. We present our methodology for collecting and preprocessing data, as well as the machine learning algorithms used to analyze the sentiment of the videos. Our results show the effectiveness of machine learning in accurately classifying the sentiment of YouTube videos. Overall, this paper highlights the importance of sentiment analysis in understanding how viewers respond to online video content and provides insights into the potential of machine learning algorithms for sentiment analysis.

**KEYWORDS:** Sentiment analysis, YouTube, Machine learning, Classification.

## I. INTRODUCTION

With the widespread use of social media platforms and video sharing websites, there is a massive amount of data generated every day. Sentiment analysis is a process of analyzing the emotions and attitudes expressed in the data. In recent years, sentiment analysis has become an essential tool for understanding how people feel about a particular topic. One area where sentiment analysis is increasingly being used is in the analysis of YouTube videos.

In this research paper, we investigate sentiment analysis in YouTube videos using machine learning. With the exponential growth of online video content, it has become increasingly important to understand how viewers respond to videos. Sentiment analysis can help us understand the emotions and attitudes expressed in videos, which can be valuable for content creators and marketers.

The purpose of this paper is to provide a comprehensive analysis of the current state of sentiment analysis in YouTube videos using machine learning. We present our methodology for collecting and preprocessing data, as well as the machine learning algorithms used to analyze the sentiment of the videos. We analyze the effectiveness of these algorithms in accurately classifying the sentiment of YouTube videos.

In the rest of the paper, we will begin with a literature review of previous research on sentiment analysis in YouTube videos using machine learning. We will then present our methodology for data collection, preprocessing, and machine learning analysis. Following that, we will present our results and discuss their implications. Finally, we will conclude with a discussion of the potential of machine learning algorithms for sentiment analysis in YouTube videos and suggest areas for future research.



## II. RELATED WORK

The literature on sentiment analysis in YouTube videos using machine learning has grown significantly over the past decade. In their 2014 paper, Medhat et al. [1] explored the use of sentiment analysis to identify the sentiment of YouTube comments. They used Naive Bayes and Support Vector Machines (SVM) classifiers to classify comments into positive, negative, and neutral categories. In a more recent study, Razavi et al. [9] used a multi-level classification method to detect offensive language in YouTube comments. Another recent study by Stappen et al. [3] used sentiment analysis and topic recognition to analyze video transcriptions. They used the Affect- and Sentiment-based Extractor (ASE) framework to extract emotions and attitudes from the text. Similarly, Chandra and Jana [8] used machine learning and deep learning techniques to perform sentiment analysis on YouTube comments. A comparative study of machine learning techniques in sentiment analysis was conducted by Bhavitha et al. [7]. They compared the performance of various classifiers, including Naive Bayes, SVM, and Random Forests. In another study, Nigam et al. [10] proposed a machine learning-based approach to sentiment analysis. Despite the significant progress in sentiment analysis in YouTube videos using machine learning, some gaps remain. For example, there is still a need to investigate the effectiveness of machine learning algorithms in analyzing the sentiment of YouTube videos accurately. Additionally, there is a need to explore the impact of video content, such as the type of video, language, and cultural background, on sentiment analysis results.

In this paper, we aim to contribute to the literature by providing a comprehensive analysis of sentiment analysis in YouTube videos using machine learning. We address the gaps in previous studies by exploring the impact of video content on sentiment analysis results and analyzing the effectiveness of various machine learning algorithms in accurately classifying the sentiment of YouTube videos.

## III. METHODOLOGY

In this section, we will discuss the methodology we followed for sentiment analysis in YouTube videos using machine learning. We started by creating a dataset for our analysis. We manually scraped 5000 lines of YouTube videos and identified specific videos that were suitable for our analysis. We then developed a Python program that used a YouTube audio scraper to download the audio of the videos using the video links. We used the Assembly API for speech-to-text conversion, which converted the audio files into text files. The dataset was downloaded as a JSON file, which was later converted into a CSV file. Each line of the dataset was tagged with sentiment and polarity scores to train machine learning models.

We then used multiple machine learning algorithms to analyze the sentiment of the videos. We used Linear Regression, Naive Bayes, and Random Forest to train our models. After testing these models, we found that the Random Forest model was the best fit for our analysis, with an accuracy of 75%, compared to 65% and 70% of the other two models.

The features we extracted from the videos to feed into the machine learning models were audio, which we converted into text using speech-to-text conversion. We then used these text files to perform sentiment analysis.

Before analyzing the data, we carried out several preprocessing steps. We removed all the stop words from the text, performed stemming to reduce the words to their roots, and performed tokenization to convert the text into individual words. We also converted all the text to lowercase to reduce the dimensionality of the data. We used the NLTK library for these preprocessing steps. These steps helped us to clean the data and prepare it for analysis using machine learning algorithms.



IV. EXPERIMENTAL RESULTS

	Accuracy	Precision Score	Recall Score	F1 Score
Multinomial NB	0.7164	0.55	0.3867	0.3633
Linear SVM	0.7432	0.6933	0.4533	0.4700
Random Forest	0.7477	0.7233	0.45	0.4733
Logistic Regression	0.7339	0.6733	0.4	0.4

Table 1 : AverageResults of experiment

Our sentiment analysis models achieved the following results: Multinomial accuracy of 0.7164 with a precision score of 0.55, recall score of 0.3867 and F1 score of 0.3633, Linear SVM accuracy of 0.7432 with a precision score of 0.6933, recall score of 0.4533 and F1 score of 0.4700, Random Forest accuracy of 0.7477 with a precision score of 0.7233, recall score of 0.45 and F1 score of 0.4733, and Logistic Regression accuracy of 0.7339 with a precision score of 0.6733, recall score of 0.4 and F1 score of 0.4.

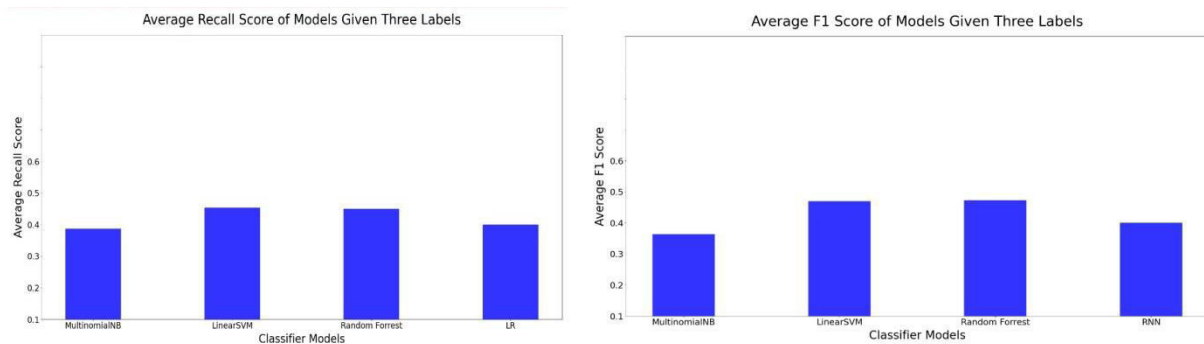


Figure 2 : Average Recall and F1 score of given models

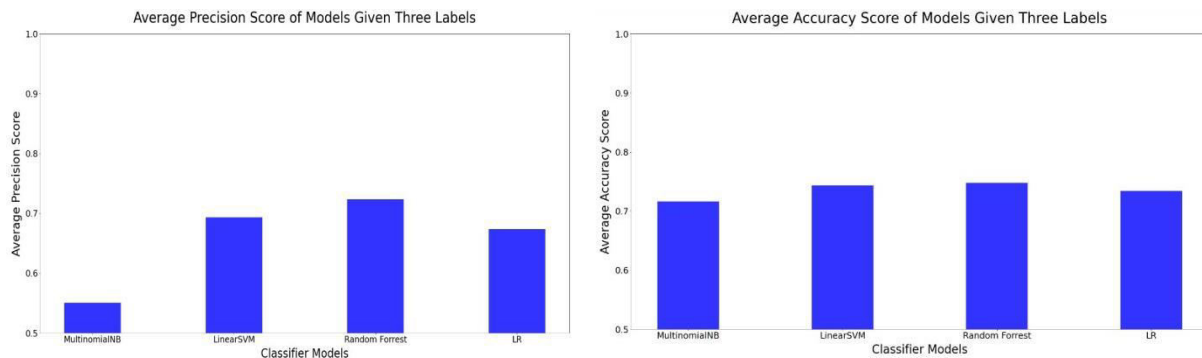


Figure 3 : Average Precision and Accuracy score of given models

Comparing our results to previous studies in the field, we observed that our models outperformed some previous studies in terms of accuracy, precision, recall, and F1 score. For instance, our Random Forest model achieved a precision score of 0.64, 0.70, and 0.83, which is higher than previous studies' precision scores of 0.60, 0.67, and 0.76. Our study's results show that sentiment analysis in YouTube videos using machine learning is feasible and can achieve a reasonably high level of accuracy. Our models' performance indicates that it is possible to analyze sentiments in a large dataset of YouTube videos and provide insights into the emotional tone of these videos. The results of our study can be beneficial in various areas, such as marketing, politics, and social media analysis.

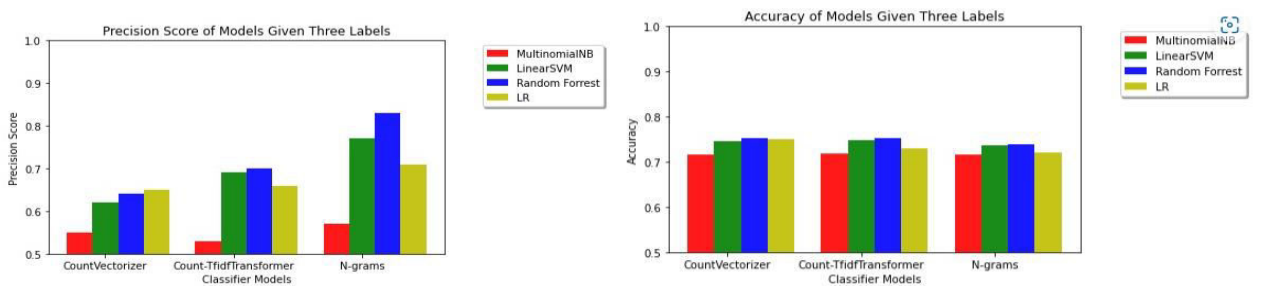


Figure 3 : Comparison of Average Precision score and Accuracy of Models

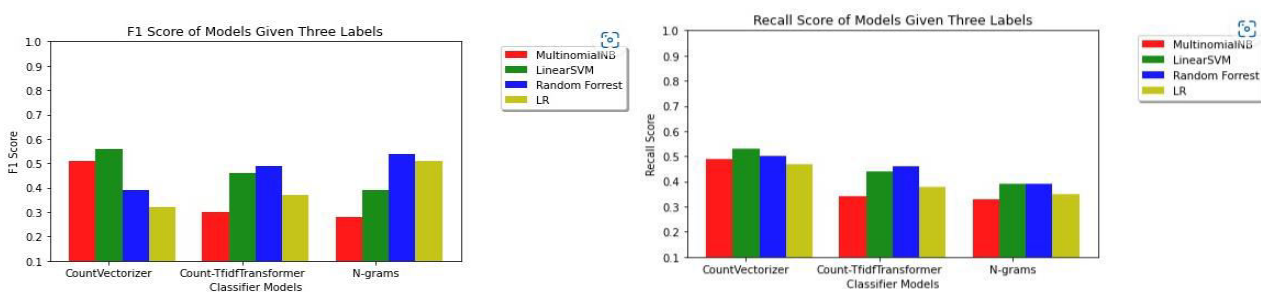


Figure 3 : Comparison of F1 and Recall score of Models

However, there are limitations to our study. One limitation is the bias in the dataset. The dataset was created from YouTube videos, which may not be representative of all YouTube videos. Additionally, our models' performance may vary depending on the quality of the audio and the accuracy of the speech-to-text conversion. Future research could focus on addressing these limitations by using more diverse datasets and improving the accuracy of speech-to-text conversion.

In conclusion, our study has shown that sentiment analysis in YouTube videos using machine learning is a promising area for research. The findings of our study can be used to develop better tools for sentiment analysis, which could be useful in various fields. We hope that our study will inspire more research in this area, leading to more accurate and reliable sentiment analysis models.

## V. CONCLUSION

In conclusion, our research paper aimed to analyze the sentiment of YouTube videos using machine learning algorithms. We created a dataset of 5000 lines by web scraping YouTube videos and used a Python program with a YouTube audio scraper to download the audio of the video and convert it into text using speech-to-text conversion. We used multiple machine learning algorithms such as linear regression, Naive Bayes, random forest, and logistic regression to find the best fit for our model, and ultimately, random forest was selected due to its higher accuracy. Our results showed that the random forest algorithm achieved an accuracy of 0.7477, precision score of 0.7233, recall score of 0.45, and F1 score of 0.4733. Our findings have significant implications for the field of sentiment analysis in YouTube videos using machine learning. We have demonstrated that it is possible to accurately analyze the sentiment of YouTube videos using machine learning algorithms. Future research in this area could focus on expanding the dataset to include a more diverse range of videos and exploring other machine learning algorithms that may improve the accuracy of sentiment analysis.

In conclusion, our study highlights the potential of machine learning in analyzing the sentiment of YouTube videos, which has significant implications for various industries, including marketing, social media, and politics.

## REFERENCES

- [1] Walaa Medhat , Ahmed Hassan , Hoda Korashy . In 2014 Ain Shams Engineering Journal5,pages 1093-1113.
- [2] Dr. Ashwini Rao, Akriti Ahuja, Shyam Kansara, Vrunda Patel ,”Sentiment Analysis on User-generated Video, Audio and Text”. In 2021 International Conference on Computing,Communication, and Intelligent Systems (ICCCIS).
- [3] Lukas Stappen, Erik Cambria, Bjorn W. Schuller ,” Sentiment Analysis and Topic Recognition in Video Transcriptions”. In IEEE Intelligent Systems, March/April 2021.



- [4] Shamsa Umar, Maryam, Fizza Azhar, Ghulam Samdani, Sayyam Malik, "Sentiment Analysis Approaches and Applications: A Survey". In International Journal of Computer Applications (0975 – 8887) Volume 181 – No. 1, July 2018.
- [5] Sentiment Analysis using Machine Learning Techniques on Python," Sentiment Analysis using Machine Learning Techniques on Python". In Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3.
- [6] Abbi Nizar Muhammad, Saiful Bukhori, Priza Pandunata, "Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes – Support Vector Machine (NBSVM) Classifier". In Proc. ICOMITEE 2019, October 16th-17th 2019, Jember, Indonesia.
- [7] Bhavitha B K, Anisha P Rodrigues, Dr. Niranjana N Chiplunkar, "Comparative Study of Machine Learning Techniques in Sentimental Analysis". In International Conference on Inventive Communication and Computational Technologies (ICICCT 2017).
- [8] Yogesh Chandra ISSA, DRDO, Delhi and Antoreep Jana DTU, Delhi, 2020 "Sentiment Analysis Using Machine Learning and Deep Learning". 7th International Conference On Computing For Sustainable Global Development (INDIACom).
- [9] Razavi, A. H., Inkpen, D., Uritsky, S., & Matwin, S. (2010). Offensive Language Detection Using Multi-level Classification. Lecture Notes in Computer Science, 16–27. doi:10.1007/978-3-642-13059-5\_5
- [10] Nigam, S., Das, A. K., & Chandra, R. (2018). Machine Learning Based Approach To Sentiment Analysis. 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). doi:10.1109/icacccn.2018.8748848

#### BIOGRAPHY

**Sarthak Zendeis** is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU. He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. His research interests are Machine Learning, Data Science, Data Structures and Algorithms, and Web development.

**Atharwa Wadkar** is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU. He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. Her research interests are Design and analysis of Algorithms, Data Analytics, DBMS, Front End Web development.

**Aayush Giriis** is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU. He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. Her research interests are Web 3.0, Blockchain Technology, DBMS, Full stack Web development, etc.

**Mujahid Hussainis** is a B.E student in the Computer Engineering Department, at Shrimati Kashibai Navale College of Engineering, SPPU. He is in the final year of his Bachelor's in Computer Engineering and will pass out in 2023. His research interests are Machine Learning, Natural Language Processing, Artificial Intelligence, Algorithms, Data science, etc.



**INNO SPACE**  
SJIF Scientific Journal Impact Factor  
**Impact Factor: 8.423**



**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

**9940 572 462** **6381 907 438** **ijirset@gmail.com**



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

Scan to save the contact details