

Rating Based on YouTube Comments

Poornima K M¹, Vibha G M², Siri K S³, Ruchitha S R⁴, Sneha S R⁵

Department of CSE JNN College of Engineering

vibhagm18122002@gmail.com

Abstract - YouTube is the most used social media platform, and it has been most popular website where users can post the videos. Opinion mining technique is used to know the sentiments expressed in YouTube comments. Focusing on diverse subjects, Naive Bayes is used to classify comments into positive, negative, and neutral sentiments. It uncovers trends in user opinions, identify preferences in teaching styles and content formats, and assess the perceived quality of educational videos. It also guides educators and content creators in optimizing educational content delivery and encourages more effective learning environment. Finally, the paper suggests reasonable approaches to achieve better results.

Keywords: Sentiment Analysis, Machine Learning, Naive Bayes.

1. INTRODUCTION

Rating based on YouTube comments is a process of understanding, extracting, and processing textual data to obtain sentiment information contained in YouTube video comment. It is additionally referred as opinion mining that means to find out or identify the positive, negative, neutral opinions, views, attitudes, impressions, emotions and feelings indicated in the text. Usually, it is used to arrive at a binary decision such as for/against, good/bad or like/dislike. Machine Learning algorithms use a training set and a test set for a classification. Training set contains information feature vectors and their corresponding class labels. By using this training set, a classification model is created which tries to classify the information feature vectors into corresponding class labels. At that point a test set is used to accept the model by predicting the class labels of unseen feature vectors. Naive Bayes is very good in classifying texts with the small amount of data. It is probabilistic classifier that makes classification using the Maximum A Posteriori (MAP) rule in a Bayesian classifier.

2. LITERATURE SURVEY

In this section, various authors have presented various Machine Learning techniques.

In [1] sentiment analysis on YouTube videos is done using the Natural Language Processing (NLP) technique and utilized the VADER (Valence Aware Dictionary for Sentiment Reasoning) lexicon to determine the sentiments of the text. use of CSV files containing comments for particular YouTube videos and a Python program that reads and further analyzes these CSV files.

In [2] Web Scraping is used to retrieve comments from the user comment section by inspecting the code of YouTube and identifying the position of the information to be extracted and have used the Waterfall method in the System Development Life Cycle (SDLC), which is one of the methods for developing application programs.

In [3] study explores the sentiments of YouTube learners by analyzing their comments on educational YouTube videos where data extraction and processing were performed using the YouTube Data API and Google Sheets and Scripts to extract YouTube comments.

In [4] research aims to analyze existing solutions for developing a sentiment lexicon specific to a domain and proposed a method applicable in this context and involves the creation of a Vietnamese sentiment lexicon (VSL) using a Term Frequency-Inverse Document Frequency (TF-IDF) algorithm.

In [5] deep learning algorithms are used specifically citing the notable successes achieved in various Natural Language Processing (NLP) tasks when compared to traditional approaches and paper introduces Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) as two prominent deep learning techniques.

In [6] paper discusses sentiment analysis on YouTube comments, covering event classification, sentiment polarity

detection using lexicons, negation detection, and Social Media Aware Phrase Detection (SMAPD). The authors propose a framework for predicting YouTube comments based on positive or negative ratings.

In [7] experiment is conducted in two phases, using N-grams features and additional text processing techniques to enhance accuracy and calculated Macro-F Score and Micro-F Score, with SVM showing the best performance followed by Naive Bayes.

In [8] compared sentiment analysis using SVM and Naive Bayes techniques and outlined the analysis into three steps: Data Collection and Preprocessing, Mining, and Result, focusing on movie user reviews. The study found that linear SVM and Naive Bayes performed well, with drama genre achieving the highest accuracy rate.

In [9] a method proposed for sentiment analysis of YouTube videos, focusing on comments, likes, and views to customize sorting of top videos using logistic regression and the YouTube API to fetch and analyze data, creating a web application that sorts videos based on sentiment, views, likes, and other parameters.

In [10] a method proposed for sentiment analysis on YouTube scrapped data. Their methodology involves extracting information from YouTube videos and user comments, storing it in a CSV file, and analyzing sentiment.

In [11] a system is developed that can perform dual-language sentiment analysis of YouTube videos in English and Roman Urdu and built a model that helps obtain reviews in both Roman Urdu and English, combines the scores, and ranks the video based on sentiment analysis in both languages.

In [12] Machine Learning algorithms are employed, specifically Naive Bayes and Support Vector Machine (SVM), for sentiment classification of Twitter reviews by using the dataset collected from Twitter, focusing on automatically detecting sentiment from the tweets.

In [13] survey of research papers related to Simple Sentiment Analysis, Complex Sentiment Analysis, and Advanced Sentiment Analysis, demonstrating that most approaches are accurate using Simple Sentiment Analysis and finally ML and deep learning techniques are widely used in Sentiment Analysis due to their high accuracy compared to other classification techniques.

In [14] analysis utilized TextBlob, a Python library for processing textual data, offering a simple API for common Natural Language Processing (NLP) tasks and concluded that YouTube comments were noted as potentially untrustworthy, as they can pertain to various domains, and the TextBlob technique provides polarity based on text without considering the context of the video.

In [15] machine learning and NLP techniques is used to sentimentally evaluate Arabic YouTube comments and Natural Language Processing (NLP) - based models are used to classify Arabic comments as positive or negative.

Authors	Title	Research focus	Remarks
Rahul Pradhan[1],2021	Extracting Sentiments from YouTube Comments.	Determine the sentiments of the text using VADER lexicon.	Limited sentiment categories.
Viny Christani M, Walda et.,al[2],2020	Comments Scraping Application for Review YouTube Content.	Extracting data through HTML tags using web scrapping.	Only opinions are used for sentiment analysis.
Ilias Chalkias, Katerina Tzafilkou et.,al[3],2023	Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering.	Explore the sentiments of YouTube by using lexicon-based sentiment analysis.	Limited comments were used for analysis.
Cong-Cuong Le, P.W.C. Prasad et.,al [4],2019	Text Classification: Naive Bayes Classifier with Sentiment Lexicon.	Model for utilizing sentiment lexicons in classifying textual data on the Internet and social media.	Creation of VSL is time consuming.
A C Nanayakkara etal. [5],2023	Sentiment Analysis of YouTube Comments Using Deep Neural Networks and Pre-Trained Word Embedding.	Deep learning techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs).	Model overfitting.
Muhammad Zubair Asghar, Shakeel Ahmad Afsana Marwat et.,al [6],2019	Sentiment Analysis Using YouTube.	Event classification, lexicon-based sentiment analysis, negation detection, and Social Media Aware Phrase Detection (SMAPD).	Limited Discussion on Practical Implementation.
Ritika Singh [7],2021	YouTube Comments Sentiment Analysis.	Aims to identify positive, negative, or neutral opinions, views, attitudes, impressions, emotions, and feelings expressed in text.	Limited Discussion on Dataset, Limited Discussion on Computational Efficiency.
Shweta Rana and Archana Singh [8],2016	Comparative Analysis of Sentiment Orientation Using SVM and Naive Bayes Techniques.	Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques.	Focus on a specific domain which may limit the generalizability of the findings to other domains.
Aditya Baravkar, Rishabh Jaiswal [9],2020	Sentimental Analysis of YouTube Videos.	The research mainly focuses on sentimental analysis of YouTube videos.	Focus on positive reviews and this may limit the applicability of the model to videos with predominantly positive sentiment in the comments.
Anusha G and Raghavendra S P [10],2022	Sentiment Analysis on YouTube Scrapped Data.	Keyword as a parameter to the URL open function, use of built-in Python libraries for coding process and results are represented using a bar chart.	It is very difficult to interpret the accuracy.

Samina Yasin and Kareem Ullah from the University of Agricultural Faisalabad, Pakistan et.,al[11],2020	Dual Language Sentiment Analysis Model for YouTube Videos Ranking Based on Machine Learning Techniques.	Developing a Dual Language Sentiment Analysis Model for YouTube ratings based on YouTube comments.	The study restricted on sentiment analysis in English and Roman Urdu.
Abdul Mohaimin Rahat, Abdul Kahir et.,al [12],2021	Sentiment Analysis Using Review Dataset.	Comparing the performance of Naive Bayes and Support Vector Machine (SVM) algorithms in sentiment analysis using a Twitter review dataset.	Specific to the characteristics of Twitter reviews.
Rawan Fahad Alhujaili and Wael M.S. Yafooz [13],2021	Sentiment Analysis for YouTube Videos with User Comments.	Implementing Sentiment Analysis methods for YouTube videos with user comments.	Model Complexity with ML techniques.
Mohd Majid Akhtar [14],2020	Sentiment Analysis on YouTube Comments: A Brief Study.	Finding the polarity of the comments.	Reliance on current sentiment dictionaries.
Dhiaa A. Musleh et.,al[15],2023	Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation.	Machine learning techniques such as Naïve Bayes (NB), Support Vector Machine (SVM), Decision Tree (DT), Random Forest, Logistic Regression, and K-Nearest-Neighbor (KNN).	Limited discussion on model interoperability.

3.CONCLUSION

This paper outlined a survey of Rating based on YouTube comments which will be done by extracting comments of YouTube videos using Scrapping. It also outlined the different approaches and techniques in various survey papers as reference points by various authors considering its advantages, and also some key challenges are discussed here. After studying various Machine Learning algorithms, it was found that Naive Bayes algorithm will produce accurate result. This survey effort will provide a better understanding of algorithms which will be used to develop the model to find the polarity of the comments.

REFERENCES

- [1] Rahul Pradhan, "Extracting Sentiments from YouTube Comments" from the Department of CEA, GLA University, published in 2021 Sixth International Conference on Image Information Processing (ICIIP) 978-1-6654-3361-7/21/\$31.00©2021IEEE|DOI:10.1109/ICIIP53038.2021.9702561 2021.PP:142-145.
- [2] Viny Christani M, Walda, and Tri Sutrisno, "Comments Scraping Application for Review YouTube Content" from the Computer Science Department, Faculty of Information Technology, Universitas Tarumanagara, published in 2020 IOP Conf. Ser.: Mater. Sci. Eng. 852 012167 2020, PP:1-6.

- [3] Ilias Chalkias, Katerina Tzafilkou, Dimitrios Karapiperis, and Christos Tjortjis, "Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering" from the School of Science & Technology, International Hellenic University, Electronics 2023, 12, 3949.

- [4] Cong-Cuong Le, P.W.C. Prasad, Abeer Alsadoon, L. Pham, and A. Elchouemi, "Text Classification: Naive Bayes Classifier with Sentiment Lexicon" published in the IAENG International Journal of Computer Science IAENG International journal of computer science 46:2, IJCS_46_2_01 ,2019 PP:1-8.

- [5] A C Nanayakkara and G A D M Thennakoon, "Sentiment Analysis of YouTube Comments Using Deep Neural Networks and Pre-Trained Word Embedding", The IUP Journal of Computer Sciences, Vol. 17, No. 3, 2023 PP:8-22.

- [6] Shakeel Ahmad Afsana Marwat, Muhammad Zubair Asghar and Fazal Masud Kundi, "Sentiment Analysis on YouTube: A Brief Survey" from the Institute of Computing and Information Technology (ICIT), Gomal University, D. I. Khan, Pakistan, Faculty of Computing and Information Technology in Rabigh (FCITR), King Abdul Aziz University

(KAU) Saudi Arabia, International Journal of Academic Research, vol 6 ,2019 PP:1-11

[7] Ritika Singh, “YouTube Comments Sentiment Analysis”, Assistant Professor in the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Ghaziabad, and Ayushka Tiwari from the Department of CSE at SRM Institute of Science and Technology, International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 05 Issue: 05 | May - 2021 PP:1-11

[8] Archana Singh and Shweta Rana, “Comparative Analysis of Sentiment Orientation Using Naive Bayes and SVM Techniques”, Amity University Uttar Pradesh ASET-IT, Noida, India, 2nd International Conference on Next Generation Computing Technologies (NGCT) Dehradun, India 14-16 October 2016 (2016), PP:106-111.

[9] Aditya Baravkar, Rishabh Jaiswal, and Jayesh Chhoriya, “Sentimental Analysis of YouTube Videos”, Department of Information Technology, Vidyalankar Institute of Technology, Mumbai International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 12 | Dec 2020 ,PP:2050-2053.

[10] Anusha G and Raghavendra S P, “Sentiment Analysis on YouTube Scrapped Data”, Department of Master of Computer Applications1 Assistant Professor, Department of Master of Computer Applications2 JNN College of Engineering, Shimoga International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), Issue 6, June 2022, ISSN (Online) 2581-9429, PP:667-671.

[11] Samina Yasin and Kareem Ullah from the University of Agricultural Faisalabad, Pakistan, Shoaib Nawaz and Zahid Aslam from the Islamia University of Bahawalpur, Pakistan, et.,al, “Dual Language Sentiment Analysis Model for YouTube Videos Ranking Based on Machine Learning Techniques”, Pakistan Journal of Engineering and Technology, PakJET Multidisciplinary | Peer Reviewed | Open Access Volume: SI, Number: 01, Pages: 213- 218, Year: 2020.

[12] Abdul Mohaimin Rahat, Abdul Kahir, and Abu Kaisar Mohammad Masum, , “Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset”, Dept. of CSE, Daffodil International University, Dhaka, Bangladesh Proceedings of the SMART–2019, IEEE Conference ID: 46866 8th International Conference on System Modeling & Advancement in Research Trends, 22nd–23rd November, 2019 College of Computing Sciences & Information Technology, Teerthanker Mahaveer University, Moradabad, India (2021),PP:266-170.

[13] Rawan Fahad Alhujaili and Wael M.S. Yafooz, “Sentiment Analysis for YouTube Videos with User Comments”, the Department of Computer Science, College of Computer Science and Engineering, Taibah University, Medina, Saudi Arabia, published in 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) | 978-1-7281-9537-7/20/\$31.00 2021 IEEE | DOI: 10.1109/ICAIS50930.2021.939604,(2021) PP:814-820.

[14] Mohd Majid Akhtar, “Sentiment Analysis on YouTube Comments: A Brief Study” (2020).

[15] Dhiaa A. Musleh, Ibrahim Alkhawaja et.,al “Arabic Sentiment Analysis of YouTube Comments: NLP-Based Machine Learning Approaches for Content Evaluation”, The Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia ,published in ISO 9001:2008 Certified Journal Volume: 07 Issue: 12 | Dec 2020 (2023).