iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57783

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Extracting Sentiments from Youtube Comments Using Deep Learning and Transformers

Ms. Julanta Leela Rachel[1], Mohammed Sawood[2], Adnan Khan[3], B. Praveen Kumar[4]
[1]*Assistant Professor*, [2, 3, 4]*Students, ISE Department, T John Institute of Technology, Visvesvaraya Technological University*

Abstract: ***YouTube is the most used social media platform, and it has been the most popular website where users can post the video. The public generally does comment, like or dislike, video-sharing on a YouTube video. Comment plays a vital role in expressing opinions and mindset, and it is used as an expression of public opinion. The massive amount of comment is generated mainly on famous channels where challenges arise to analyse public opinion or behaviour regarding that particular video. This project proposes sentiment analysis on YouTube video by Natural Language Processing (NLP) technique along with Deep Learning techniques Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) models. Sentiment analysis is when comprehension, citation, and processing of text-based data is done, and it directly converts it into sentiment information. This analysis help users to get the report of their YouTube video. The output of this analysis gives the classification of sentiment analysis, as Positive, negative, or neutral.***
Keywords: Sentiment Analysis, Recurrent Neural Network, Gated Recurrent Unit, Natural Language Processing.

I. INTRODUCTION

Sentiment analysis on YouTube is becoming increasingly important for businesses and content creators. It allows them to understand how their audience is responding to their content and make informed decisions about future content creation or marketing strategies. Sentiment analysis can also help businesses and content creators identify areas for improvement and address negative feedback to enhance their brand reputation. With help of Metadata (comments) we have the potential to correctly find popularity of the video. Since the invention of computer linguistics, text mining, and sentiment analysis, determining the polarity of words in a certain context has been possible. Machine learning-based methods use a supervised learning mechanism for sentiment. This indicates the polarity information (i.e., positive, negative, and neutral) and an assigned numeric value to score how positive or negative a given word is. The lexicon-based methods, NLP methods are improved method for sentiment analysis. One of the biggest challenges in sentiment analysis on YouTube is the large volume of comments that need to be analysed. Moreover, the informal nature of YouTube comments and the use of slang and jargon can make it difficult for machine learning algorithms to accurately interpret the sentiment of comments. Therefore, natural language processing techniques and sentiment lexicons need to be carefully selected to achieve accurate sentiment analysis on YouTube comments.

II. LITERATURE REVIEW

[1] The study on sentiment analysis for YouTube comments using machine learning techniques published on May 2023. The paper evaluates three supervised machine learning classifiers: Decision Tree, K-Nearest Neighbours and Support Vector Machine, to see how accurately they predict the public sentiment in YouTube reviews. The paper also compares the results with other existing approaches and shows that the machine learning-based model outperforms alternative methods. The paper reviews the literature on sentiment analysis and machine learning, and proposes a methodology that involves data pre-processing, feature selection, ensemble learning, and model evaluation. The paper concludes by suggesting future directions for research, such as exploring other machine learning techniques and software metrics.

[2] This study aims to determine if "sentiment analysis on YouTube comments can be useful for predicting the like proportion of YouTube videos". The authors use five supervised machine learning classifiers: logistic regression, support vector machine, stochastic gradient descent classifier, multinomial naive Bayes, and complement naive Bayes algorithms to classify YouTube comments as positive, negative, or neutral. They use four different formulas to predict the like proportion based on the number of comments in each sentiment category. The authors find that training the classifiers on YouTube comments only outperforms training on tweets or a combination of tweets and comments. They also find that attributing all neutral comments to likes gives the highest correlation and the lowest error, but this may be biased by the high average like proportion of the testing dataset.

©IJRASET: All Rights are Reserved

The authors conclude that there is some positive correlation between the comment sentiment and the like proportion, but the prediction accuracy is not very high and needs improvement.

[3] The research article named "A Deep Neural Network-Based Approach for Sentiment Analysis of Movie Reviews" employs three main methodologies for sentiment analysis: lexicon-based techniques, machine learning and hybrid-based methods. Lexicon-based techniques utilize dictionary-based or corpus-based approaches, with the former employing dictionary terms like SentiWordNet and WordNet, and the latter using statistical analysis. Machine learning, including supervised and unsupervised learning is utilized to automatically learn from labelled or unlabelled data, while the hybrid approach combines both lexicon-based and machine learning methods. The study proposes a seven-layer deep neural network model for sentiment analysis of movie reviews, leveraging techniques such as Word2Vec for word vector conversion and convolutional layers for feature extraction. The model's performance is evaluated using accuracy as the primary metric, achieving notable results in classifying the sentiment of movie reviews, with specific experiments demonstrating improved accuracy when processing longer or average-length reviews compared to shorter ones. The findings highlight the potential of the deep neural network-based approach for effectively analysing sentiment in movie reviews, offering valuable insights for applications in understanding user opinions and preferences in the film industry.

[4] The authors exhibited a hierarchical framework concentrating on aspect-specific sentiment analysis. They developed novel d-dimensional vector representations for words to extract labels at the phrase level. They introduced a deep learning framework involving dealing with feature representations of parses of the sentences, which contribute to an objective function to be determined. For this, multi-vector RNN and recursive neural tensor network are compared alongside vanilla RNN. These are attached to the aspect and sentiment labels using their joint multi-aspect sentiment model. They compared their model for single and joint aspect sentiment pair detection and compared it against multiclass SVMs and naive Bayes classifiers.

[5] The authors developed a method that operates at the character level and uses a bi-LSTM (bi-Long Short Term Memory) in addition to a conditional random field classifier (Bi-LSTM-CRF) and an aspect-based LSTM to classify the polarity in clauses of text. These are demonstrated on the dataset of Arabic Hotels' reviews. They focused on aspect sentiment polarity identification after the completion of OTE, and their results show that their approaches outperformed previous research methodologies.

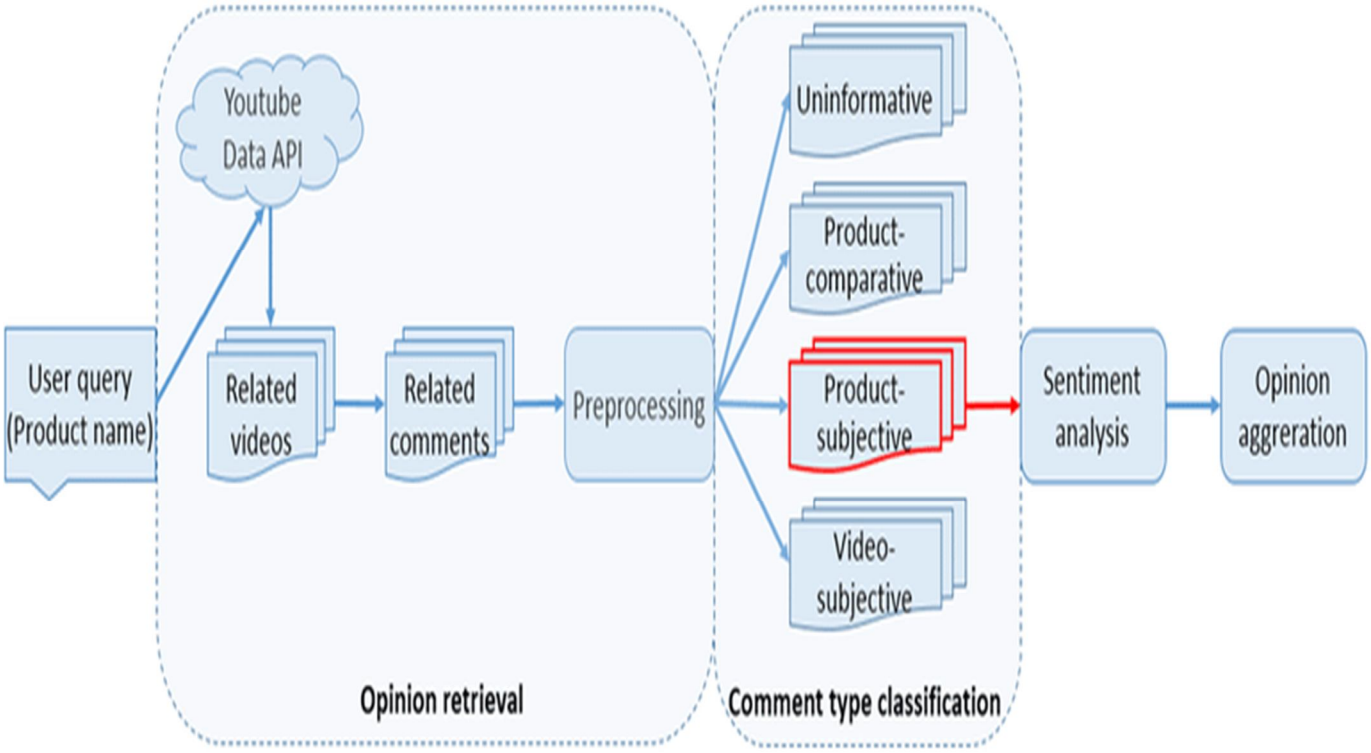
TABLE 1.1

SI.NO	References taken from other research/work		
	Title	Author's	Work
6	Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism.	Zeng D, Dai Y, Li F, Wang J, Sangaiah AK	GLRC using linguistic rules to enhance the results. Data were embedded with two linguistic regularizers - ACR and GCR.
7	Attentional recurrent neural networks for sentence classification.	Kumar A, Rastogi R	RNN, Bi-RNN, GRU, LSTM, Higher accuracy was achieved (accuracy: 89.60% on MPQA dataset).
8	Multimodal sentiment analysis using hierarchical fusion with context modelling.	Majumder N, Hazarika D, Gelbukh A, Cambria E, Poria S	Achieved an accuracy of about 1–2% more than the existing ones. Textual modality was 21% better compared to the audio modality (accuracy: IEMOCAP: 79.6%).
9	Target-dependent sentiment analysis of tweets using bidirectional gated recurrent neural networks.	Jabreel M, Hassan F, Moreno A	Td-Bi-GRU got an accuracy 4% higher than what it was before in the previous basic GRU model (accuracy: 72.25%).
10	Bidirectional-GRU based on attention mechanism for aspect-level sentiment analysis.	Penghua Z, Dingyi Z	ABAE-Bi-GRU model achieves outstanding performance and results in greater accuracy at datasets and has further improvements compared to previous models (accuracy: restaurant dataset: 81.2%)

III. METHODOLOGY

In the digital age, user-generated content on platforms like YouTube has exploded, offering an immense wealth of opinions and sentiments. Our project, titled "Extracting Sentiments from YouTube Comments using Deep Learning and Transformers," endeavours to harness this treasure trove of data for sentiment analysis. To achieve this, we employ cutting-edge techniques, including Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) models, which have demonstrated remarkable performance in natural language processing tasks. Our primary dataset is the Review Sentiment dataset, a well curated collection of text reviews with associated sentiment labels. Leveraging the vastness and diversity of YouTube comments, we adapt these DL models to decipher nuanced sentiments expressed in online discussions. This research not only contributes to the field of sentiment analysis but also addresses the challenges posed by user-generated content in the YouTube ecosystem. Our findings hold the potential to inform content creators, marketers, and platform moderators about audience sentiment, thereby facilitating more informed decision making in the digital realm.

IV. PROPOSED ARCHITECTURE



V. CONCLUSION

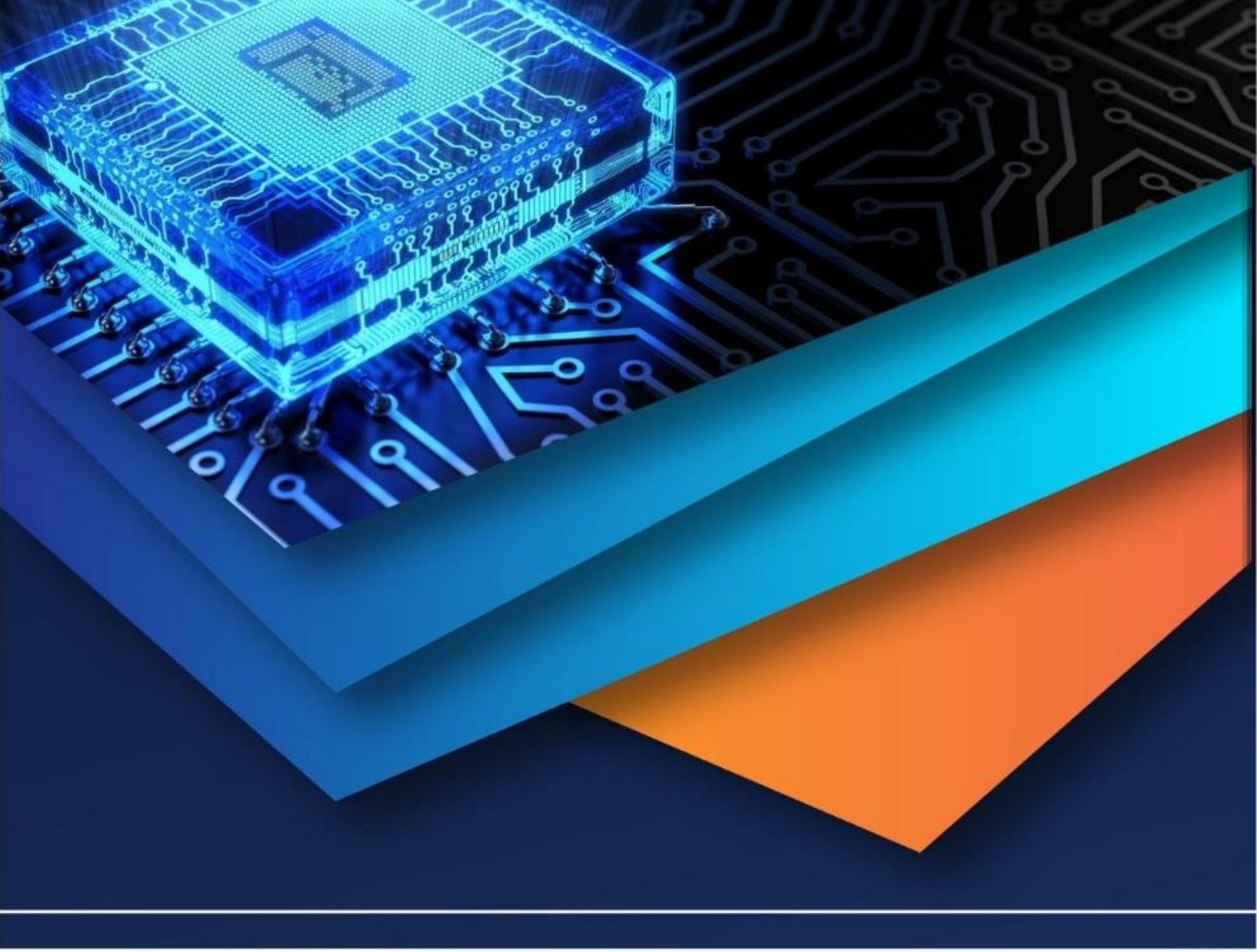
After analysing all these studies, it can be concluded that deep learning methods can obtain better results than non-deep learning methods for sentiment analysis. This project aims to summarize, detect, and classify different types of user-generated text, including tweets, comments, and other forms of text. Machine learning techniques will be used to classify the text into various types. The insights obtained from the analysis will be presented in an interactive dashboard for easy interpretation. Proper citation and referencing will be used throughout the project to ensure originality.

REFERENCES

[1] Sainath Pichad, Sunit Kamble, Rohan Kalamb, Sumit Chavan, “Analysing Sentiments for YouTube Comments using Machine Learning”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 11 Issue V, May 2023.
[2] Isac Lorentz, Gurjiwan Singh, “Sentiment Analysis on YouTube Comments to Predict YouTube Video Like Proportions” KTH Royal Institute of Technology, 2021.
[3] Kifayat Ullah, Anwar Rashad, Muzammil Khan, Yazeed Ghadi, Hanan Aljuaid, and Zubair Nawaz, “A Deep Neural Network-Based Approach for Sentiment Analysis of Movie Reviews”, Hindawi Complexity Volume 2022, Article ID 5217491, 9 pages.

[4] Lakkaraju H, Socher R, Manning C, "Aspect specific sentiment analysis using hierarchical deep learning" In: NIPS Workshop on deep learning and representation learning; 2014.
[5] Al-Smadi M, Talafha B, Al-Ayyoub M, Jararweh Y, "Using long short-term memory deep neural networks for aspect- based sentiment analysis of Arabic reviews" Int J Mach Learn Cybern. 2019;10(8):2163–75.
[6] Zeng D, Dai Y, Li F, Wang J, Sangaiah AK, "Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism" J Intell Fuzzy Syst. 2019;36:3971–80
[7] Kumar A, Rastogi R, "Attentional recurrent neural networks for sentence classification." In Innovations in infrastructure, Springer; 2019. pp. 549–59.
[8] Majumder N, Hazarika D, Gelbukh A, Cambria E, Poria S, "Multimodal sentiment analysis using hierarchical fusion with context modelling.", Knowl-Based Syst. 2018;161:124–33.
[9] Jabreel M, Hassan F, Moreno A, "Target-dependent sentiment analysis of tweets using bidirectional gated recurrent neural networks" In: Advances in hybridization of intelligent methods. Cham: Springer; 2018. p. 39–55.
[10] Penghua Z, Dingyi Z, "Bidirectional-GRU based on attention mechanism for aspect-level sentiment analysis", In: Proceedings of the 2019 11th international conference on machine learning and computing. ACM; 2019. p. 86–90.
[11] Amna Hafiz, Uzma Raja, and Muhammad Farhan, "Sentiment analysis using deep learning: A review" In 2020 IEEE 12th International Conference on Quality, Reliability, Infocom Technology and Industrial Application (ICQRITIA), pp. 320-324. IEEE, 2020.
[12] Jiashen Liu, Jia Liu, Xinyi Wang, and Jian Zhang. "Sentiment analysis on social media: A survey" IEEE Transactions on Computational Social Systems, 7(3), pp. 682-705, 2020.
[13] Xiaoxu Liu, Jianquan Liu, Xianping Tao, and Yuexiang Yang, "A review of sentiment analysis research based on deep Learning" Journal of Ambient Intelligence and Humanized Computing,12(2), pp. 1475-1489, 2021.
[14] Khaled Abdalgader, Aysha Al Shibli "Experimental Results on Customer Reviews Using Lexicon-Based Word Polarity Identification Method" in IEEE Access (Volume: 8), October 2020.
[15] Adnan Ishaq, Sohail Asghar, Saira Andleeb Gillani "Aspect-based sentiment analysis using a hybridized approach based on CNN and GA" in IEEE Access (Volume: 8), July 2020.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)