OPTIMIZING SENTIMENT ANALYSIS IN EDUCATIONAL YOUTUBE VIDEOS: A COMPARATIVE STUDY OF ROBERTA AND MULTINOMIAL NAIVE BAYES

Ulima Inas Shabrina[1], Muhammad Iskandar Java[2], and Siti Rochimah[3]
[1, 2,3]Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
e-mail: 6025231046@student.its.ac.id[1], 6025231013@student.its.ac.id[2], siti@its.ac.id[3]

ABSTRAK

YouTube has evolved into a globally influential platform, engaging over 2.1 billion users worldwide and serving as a prominent medium for sharing, consuming, and creating diverse video content. Particularly popular among younger demographics, YouTube stands as a multifaceted hub spanning various genres and has significantly impacted education by providing extensive educational materials, fostering independent learning, and supporting a wealth of educational resources. This research conducts an in-depth investigation into sentiment analysis specifically within the context of educational YouTube videos. Leveraging advanced machine learning techniques, notably RoBERTa, this research conducts a comparative analysis with Multinomial Naive Bayes (MNB). The primary focus is on exploring RoBERTa's adaptability and performance across a spectrum of educational video content, revealing its commendable accuracy of 91.21%, surpassing MNB's accuracy of 79.59%. However, it is observed that RoBERTa's performance is notably affected by smaller datasets, highlighting the critical importance of ample and diverse training data for achieving optimal results. These findings highlight the pivotal role of dataset characteristics and size in developing robust sentiment analysis models, especially with advanced natural language processing methods like RoBERTa.

***Keywords**: Sentiment Analysis, Educational Multimedia, RoBERTa, Multinomial Naive Bayes*

MENGOPTIMALKAN ANALISIS SENTIMEN PADA VIDEO EDUKASI YOUTUBE: STUDI PERBANDINGAN ROBERTA DAN MULTINOMIAL NAIVE BAYES

Ulima Inas Shabrina[1], Muhammad Iskandar Java[2], dan Siti Rochimah[3]
[1,2,3]Institut Teknologi Sepuluh Nopember
Surabaya
e-mail: 6025231046@student.its.ac.id[1], 6025231013@student.its.ac.id[2], siti@its.ac.id[3]

ABSTRAK

YouTube telah berkembang menjadi platform yang berpengaruh secara global, melibatkan lebih dari 2.1 miliar pengguna di seluruh dunia dan menjadi medium yang menonjol untuk berbagi, mengonsumsi, dan menciptakan beragam konten video. Terutama populer di kalangan demografis yang lebih muda, YouTube menjadi pusat multifaset yang meliputi berbagai genre dan telah berdampak signifikan pada pendidikan dengan menyediakan materi edukatif yang luas, mendorong pembelajaran mandiri, dan mendukung banyak sumber daya pendidikan. Studi ini melakukan investigasi mendalam terhadap analisis sentimen khususnya dalam konteks video YouTube yang bersifat edukatif. Dengan memanfaatkan teknik-teknik machine learning canggih, terutama RoBERTa, penelitian ini melakukan analisis komparatif dengan Multinomial Naive Bayes (MNB). Fokus utamanya adalah untuk mengeksplorasi adaptabilitas dan kinerja RoBERTa dalam rentang konten video edukatif, mengungkapkan akurasinya yang memuaskan sebesar 91.21%, melampaui akurasi MNB sebesar 79.59%. Namun, diamati bahwa kinerja RoBERTa secara signifikan dipengaruhi oleh dataset yang lebih kecil, menyoroti pentingnya data pelatihan yang cukup dan beragam untuk mencapai hasil yang optimal. Temuan ini menyoroti peran penting karakteristik dan ukuran dataset dalam mengembangkan model analisis sentimen yang tangguh, terutama dengan metode pemrosesan bahasa alami canggih seperti RoBERTa.

***Keywords**: Sentiment Analysis, Educational Multimedia, RoBERTa, Multinomial Naive Bayes*

I. INTRODUCTION

IN 2020, YouTube boasted over 2.1 billion users, with more than 95% of the global Internet population actively engaging with the platform across 76 different languages in over 88 countries [1]. This platform, highly favored by the youth demographic, captivates users for hours as they consume online videos, interact with fellow users, and sometimes produce their own content [1]. Furthermore, YouTube offers a vast array of video categories, spanning education, entertainment, and informational content. Users can seamlessly view, share, upload, and engage through comments on diverse videos, rendering it a versatile platform for both creators and viewers [2]. This global appeal extends YouTube's immense popularity beyond borders and continents, solidifying its status as a truly global hub for video streaming and content creation [1].

YouTube is widely used for learning and sharing educational content. Universities and educational institutions employ YouTube as an additional teaching resource, acknowledging the significance of videos and visual elements in facilitating effective learning [3]. Both students and learners have the opportunity to explore a diverse selection of educational videos covering a multitude of subjects, which proves to be a valuable asset for independent learning.

Comments left on videos provide additional information, perspectives, and viewpoints regarding the content [4]. Viewers have the opportunity to peruse these comments, engaging with them to enhance their comprehension of the video's quality and its applicability to their requirements. In the realm of educational videos, individuals can depend on comments and other cues to evaluate the content's caliber. For instance, an examination of YouTube's educational videos investigated how various video attributes correlated with user assessments of the content [4]. Sentiment analysis algorithms are used to computationally analyze the opinions, sentiments, and subjectivity of text data [5]. These algorithms can assist in gauging users' opinions by analyzing the sentiment conveyed in their comments. Naive Bayes, a supervised machine learning technique, can be used for sentiment analysis and is frequently assessed alongside other classifiers like K-Nearest Neighbors (K-NN), Support Vector Machines (SVM), and Decision Trees to gauge its effectiveness.

Many studies have been conducted on user interaction patterns on YouTube. This interaction takes the form of comments, which are typically used to express feelings and provide emotional support, impressions, or advice in response to videos [6]. User behavior is widely utilized for various purposes, such as marketing objectives, where companies can use the information to target their messages and products [7], in politics to support campaign strategies, to evaluate government services [8], and more. Therefore, YouTube serves not only as entertainment but also for branding, education, sports, health, documentaries, and so on [9]. This research aims to delve into the workings of each existing method so that future research can use the appropriate method with accurate results without wasting much time. Another study conducted by [10] achieved an accuracy rate of 94% for the naive Bayes algorithm compared to the SVM algorithm with an accuracy rate of 75.50%. This indicates that the naive Bayes algorithm is superior to the SVM algorithm in classifying Twitter comments. In the context of educational YouTube channels, it is essential to gain insights into its reach, understand its role in education as a valuable resource for both learners and educators and recognize the significance of user comments in assessing video content quality and relevance. These preliminary considerations set the stage for conducting sentiment analysis on specific comments from various educational YouTube channels.

In this research, we propose a transfer learning approach aimed at bolstering sentiment analysis within the context of educational content on YouTube. Specifically, we leverage RoBERTa (Robustly Optimized BERT Approach), a pre-trained model initially calibrated on Twitter datasets, as the foundation for our approach. This approach helps the model adapt to the unique features of educational videos without having to learn from scratch. To adapt this model to the unique characteristics of educational video content, we undertake a process of retraining using fresh datasets sourced from diverse channels on YouTube. Our objective is to explore and evaluate the adaptability and performance of the RoBERTa-based model across a spectrum of educational videos from various channels. Through a comprehensive analysis, we aim to discern how well our proposed approach performs in discerning sentiments within the nuanced context of different educational video content, ultimately contributing valuable insights to the field of sentiment analysis in educational multimedia.

II. RELATED WORKS

This research [11] discusses the development of a solution for automating sentiment analysis of YouTube video comments. It aims to help content creators quickly understand viewer preferences without manual comment reading. The data used comes from comments on various YouTube channels with different themes such as daily vlogs, culinary videos, unboxing, hack/DIY, and music covers. The analysis results indicate that daily vlog content is preferred, with a positive sentiment of 84% and a negative sentiment of 16%. The model using the SVM algorithm achieves an accuracy of 86%, precision of 87%, recall of 99%, and an f1-score of 100%. This suggests that the model effectively classifies comment sentiments, assisting content creators in automating viewer preference understanding [11].

This research [12] discovered that conducting sentiment analysis on comments posted on YouTube videos involves automatically processing textual data to extract sentiment information contained within each comment. Employing text mining methods emerges as a favorable approach to decipher the essence of individual comments. The categorization of comments into positive or negative holds significant importance for YouTube users, aiding in the assessment of the content's meaningfulness based on user opinions. While Naïve Bayes and Support Vector Machine serve as fundamental approaches in text-related tasks, their performance varies notably across different variations, features, and quantities of data collected. Naïve Bayes demonstrates proficiency in classifying texts with limited data and shorter document snippets, whereas Support Vector Machines excel in categorizing texts with relatively larger data sets or full-length documents. Combining Naïve Bayes and Support Vector Machine methods yields heightened accuracy and stronger performance, particularly when using a 7:3 data split (70% training data and 30% testing data). This combination exhibits the highest performance in testing, showcasing precision of 91%, recall of 83%, and an f1-score of 87%.

This research [13] learned that film reviews, inherently subjective in nature, pose a challenge for enthusiasts trying to assess movies in line with their preferences. To address this issue, the study turned to sentiment analysis, also referred to as opinion mining, which systematically assigns emotional labels to textual content to distinguish positive or negative sentiments. The Naive Bayes method was chosen for its ability to classify data based on class probability computation. Comparing data prepared with and without lemmatization, the study found that the most effective model emerged from un-lemmatized data, utilizing a 500-vector size and Naive Bayes classification, resulting in an accuracy of 78.96 percent and an f1-score of 78.81 percent. Additionally, the research uncovered that variations in vector size significantly impact the system's predictive capacity, indicating lower precision and recall outcomes with a vector size of 300 compared to the more robust results obtained with a vector size of 500. These insights highlight the crucial role of vector size in enhancing sentiment analysis models for discerning sentiments in film reviews.

This research [14] discusses the importance of identifying user opinions to create appealing content. Its purpose is to evaluate the quality of video content based on user opinions. Data is gathered from comments on YouTube videos in the education category using crawling API techniques on the "Kok Bisa?" channel. The study employs the Naïve Bayes - Support Vector Machine (NBSVM) method with a Binary Classification approach. Research results show that the combination of Naïve Bayes and Support Vector Machine yields a good level of accuracy, with a precision of 91%, recall of 83%, and an average score of 87% in performance tests. This suggests that the method is effective in assessing content quality based on user opinions, particularly for educational content on the "Kok Bisa?" channel [12].

In a research conducted by Xu *et al.*, a continuous naive Bayes learning approach was employed to perform sentiment classification on e-commerce product reviews. This innovative continuous naive Bayes learning framework represents an enhancement over the conventional naive Bayes method by retaining knowledge acquired from previous domains and facilitating learning in new domains. This framework effectively tackles the challenges related to computational efficiency and continual learning in the context of sentiment classification. As a result of this approach, high accuracy is attained, and it demonstrates exceptional adaptability to diverse domains [15].

Alkaff *et al.* conducted research focused on sentiment analysis of Indonesian movie trailers found on YouTube. This analysis involves the utilization of the Delta TF-IDF word weighting technique in combination with the Support Vector Machine (SVM) algorithm. The study categorizes movie comments into four prevalent genres: action, romance, comedy, and horror. It aims to determine whether there are notable distinctions in the expression of positive and negative sentiments corresponding to each genre. The investigation assesses the performance of the Delta TF-IDF and SVM model by comparing it to other established classification methods, namely Naïve Bayes and Logistic Regression when classifying sentiment across all movie trailer genres. The study's findings indicate that, for classifying sentiment within a specific genre, Logistic Regression and Naïve Bayes outperform the SVM model. However, the SVM model excels in sentiment analysis when applied to a more general genre [16].

From the previous studies, we can conclude that SVM and Naive Bayes are commonly used for sentiment analysis from various domains. However, it is crucial to acknowledge certain limitations associated with traditional approaches. Naive Bayes, for instance, operates on the assumption of word or token independence, which diverges from the reality of natural language where words or phrases often exhibit interdependence [5]. Similarly, Support Vector Machines (SVMs), designed primarily as binary classifiers, may not demonstrate optimal performance when adapted for multi-label classification, particularly in comparison to algorithms explicitly tailored for such scenarios [17]. These limitations, marked by the reliance on manually crafted features and the inability to capture intricate language relationships, underscore the necessity for more sophisticated methodologies.

To address these issues, our research introduces a novel approach leveraging transfer learning based on the RoBERTa model for the classification of educational comments from various channels on YouTube. RoBERTa, a variant of the BERT model, serves as a pre-trained language representation model renowned for its exceptional performance across a wide spectrum of natural language processing tasks [18]. This strategic integration aims to overcome the limitations posed by Naive Bayes and SVMs, offering a more advanced and effective solution for sentiment analysis in the specific context of YouTube educational content.

III. Methods

The stages involved in this research can be seen in Fig 1.

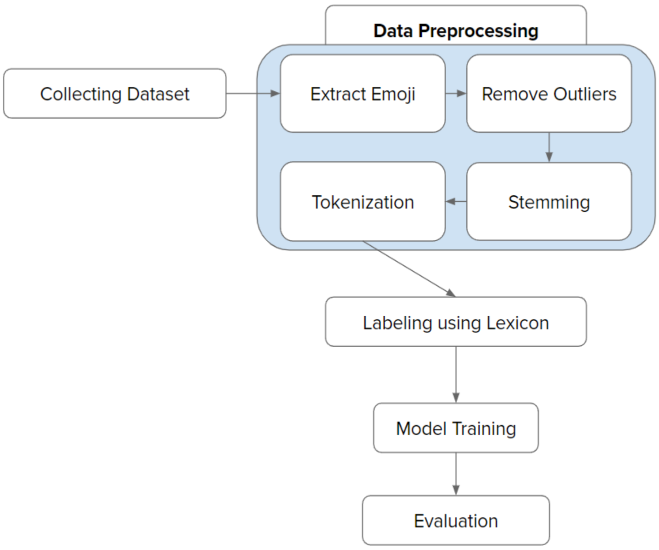

Fig. 1. Process Method

A. Web Youtube API Scrapping

The Web Scraping technique is used to extract comment data from YouTube. Web scraping is a data retrieval method that serves as a dataset collector and is the first step in this research. Scraping is performed using the YouTube API. Reference to the specific channels involved is detailed in Table I, which includes comments gathered from five YouTube channels: Mindvalley Talks, TedX, Kurzgesagt, NeetCode, and Be Smart.

Table I Total Dataset from YouTube Scrapping

Video	Chanel	Total
Speed Learning: Learn In Half The Time | Jim Kwik	Mindvalley Talks	821
Unleash Your Super Brain To Learn Faster	Mindvalley Talks	3.914
Change Your Frequency, Change Your Reality	Mindvalley Talks	2.332
The power of vulnerability | Brené Brown	TedX	3377
How not to take things personally? | Frederik Imbo | TEDxMechelen	TedX	9609
How a 13 year old changed 'Impossible' to 'I'm Possible' | Sparsh Shah | TEDxGateway	TedX	5324
Why You Are Lonely and How to Make Friends	Kurzgesagt	10.281
The Most Horrible Parasite: Brain Eating Amoeba	Kurzgesagt	12.064
What Actually Happens When You Are Sick?	Kurzgesagt	5372
My Brain after 569 Leetcode Problems	NeetCode	553
Two Sum - Leetcode 1 - HashMap - Python	NeetCode	156
GPT-4 is OVERHYPED	NeetCode	244
How An Igloo Keeps You Warm	Be Smart	4270
What Is Fire?	Be Smart	5560
Why Do We Itch?	Be Smart	7417

B. Pre-processing

Preprocessing data involves cleaning the data so that it can be utilized for the subsequent stages. The preprocessing steps conducted in This research include case folding, cleansing, stemming, stopwords, and tokenization.

- Cleansing is the process of removing unnecessary text from the document, such as hashtags, HTML, emoticons, URLs, and mentions.
- Stemming aims to transform words in the text document into their root forms.
- Tokenization is the process of splitting each word in a sentence using quotation marks (") as separators [19].

C. Labelling

The labeling process is done using a lexicon language dictionary by first clearing the default lexicon language dictionary, and then populating the lexicon dictionary using an English sentiment lexicon.

D. BERT

BERT is a language representation model designed to train a two-way representation of unlabeled text by conditioning left and proper contexts at all layers. Bidirectional Encoder Representations from Transformers (BERT) optimize Masked Language Model (MLM) and Next Sentence Prediction (NSP) in the Pre-Trained process. The masked Language Model (MLM) is a process model to predict the words that will appear from the previous comments. Next Sentence Prediction (NSP) is a binary classification loss that functions to indicate two words that follow each other in a text [14]. BERT was the first finetuning-based representation model to outperform multiple-task architectures [18]. The training analysis of the BERT method explores and quantifies the options that are important for training the BERT model while maintaining the architectural model. This starts from the training of the BERT model with the same configuration as BERT base (training ^{2}L = 12, H= 768, A = 12, 110M params) [20].

E. RoBERTa

RoBERTa, an enhanced version of BERT, refines the Pre-Trained approach by optimizing the system to forecast text sections in annotated languages using a comprehensive dataset of 160GB. It stems from the replicated research of BERT pre-training, emphasizing the impact of hyperparameters and training dataset size. This model excels in predicting text portions in annotated languages and comprises two pre-trained variants: RoBERTa large and RoBERTa base, both structured based on the BERT architecture [20].

The process involves a detailed analysis employing pre-training and fine-tuning procedures. Pre-training aims at establishing a pre-trained RoBERTa model, initially configuring it by defining the utilized hyperparameters, acquired tokens, and the training dataset. Fine-tuning follows pre-training, utilizing the pre-trained model to execute the masking process. Subsequently, data preprocessing is conducted to compile data for training, testing, and validation. This phase is finalized by configuring the model using the hyperparameters specified for the fine-tuning process [21].

RoBERTa's enhanced performance hinges on four crucial elements [20]. Initially, it employs the Dynamic Masking architecture during training, generating a mask pattern for each sequence entered into the model. This proves pivotal, particularly during extensive pre-training with more steps or richer, more meaningful data. Secondly, RoBERTa discards the Next Sentence Prediction (NSP) task, which prompts users to assess the relationship between two sentences. This omission is based on Liu's research (2009), indicating that employing just one sentence adversely affected downstream task performance. Consequently, RoBERTa utilizes entire sentences without NSP to improve overall performance [20]. Thirdly, leveraging a large number of batches during training significantly accelerates optimization speed and final task performance. RoBERTa substantially scales up the batch size from 256 BERT sequences to 8,000 sequences within its architecture, enhancing efficiency. Fourth, RoBERTa employs robust Byte-Pair Encoding (BEP), a fusion of character and word representation facilitating the handling of extensive vocabularies. In its BEP implementation, RoBERTa utilizes bytes as the foundation for subword units, allowing the exploration of subword vocabulary units of up to 50,000 [20]. These optimizations result in several advantages, including increased periods with more batches of data and the removal of the NSP task, enabling RoBERTa to utilize Dynamic Masking, thus enhancing its performance. RoBERTa consists of two pre-trained models, RoBERTa large and RoBERTa base(L), both designed following the BERT architecture [14]. RoBERTa large and RoBERTa base employs a training model configured extensively within the BERT architecture (training ^{2}L = 24, H = 1024, A = 16, 355M params). The specific hyperparameters utilized in RoBERTa's pre-training and refinement processes are detailed in the aforementioned study [20].

F. Naïve Bayes Classifier

The method of naive Bayes classification relies on Bayesian principles and employs statistical techniques to forecast the probability of belonging to specific classes. It operates under the naive Bayes assumption, suggesting that the attribute value of one class doesn't influence the attribute value of another class (known as class-conditional independence) [22] [23]. Consequently, the overall probability distribution is computed by multiplying the probability distributions of individual data points, as shown in the equation below:

$$p(d|c) = \prod_{i=1}^{m} P(w_i|c) \tag{1}$$

$$P(w_i|c) = \frac{\sum_{j=1}^{n} \delta(w_{ji}, w_i)\delta(c_j, c) + 1}{\sum_{j=1}^{n} \delta(c_j, c) + n_i} \tag{2}$$

In this context, n represents the count of training instances, n_i signifies the number of values associated with the ith attribute, c_j denotes the class label assigned to the jth training instance, w_{ji} represents the ith attribute value of the jth training instance, and $\delta(\cdot)$ is a binary function. The binary function evaluates to one if its two parameters are identical and zero otherwise [24].

IV. RESULT AND DISCUSSION

In our research, we employed the proposed methodology alongside multinomial Naive Bayes to classify the predicted sentiment of each comment into categories of neutral, positive, or negative. Subsequently, the predicted labels from the test dataset were utilized to assess and compare the performance of both methods. It is crucial to note that the datasets utilized in this research exhibit variations both in terms of quantity and distribution, adding an extra layer of complexity to the evaluation process. For the hyperparameters used in the RoBERTa models, we set the "adam_beta1" parameter to default value 0.9. This adjustment adjusts the weight given to past gradients in the Adam optimizer, making the optimization process more stable. Furthermore, we use a learning rate value of 25e-6 with 10 training epochs. For the multinomial Naïve Bayes models, we use default parameters. Therefore, this comprehensive evaluation of sentiment prediction not only involved the classification into distinct sentiment categories but also included a thorough examination of the effectiveness and accuracy of each method in handling the diverse dataset present in the educational YouTube video comments.

In order to comprehensively extract sentiment from comments spanning various videos and channels, we established specific criteria to ensure data quality and relevance. Firstly, comments had to be composed in English, facilitating consistent sentiment analysis. Secondly, the selected videos had to belong to the educational genre, aligning with our analytical focus. Furthermore, we imposed a threshold of at least 100 comments per video, aiming to gather a sufficiently robust dataset for sentiment analysis. However, obtaining comments that met our criteria proved to be quite challenging. We encountered difficulty in finding educational videos on YouTube that garnered more than 100 comments. To address this challenge, we strategically explored popular YouTube channels known for their educational content. Our exploration led us to channels such as Mindvalley Talks, TedX, Kurzgesagt, NeetCode, and Be Smart, which consistently produce insightful and engaging educational videos.

In the analysis of all videos from the different channels, RoBERTa consistently outperformed the traditional Naive Bayes approach in terms of accuracy, F1 score, and precision. For instance, in the video titled "Why You Are Lonely and How to Make Friends," RoBERTa achieved an accuracy of 91.34%, an F1 score of 91.34%, and a precision of 91.34%, while the performance of Naive Bayes was comparatively less proficient, with an accuracy of 76.50%, an F1 score of 67.00%, and precision of 75.00%. This trend continued across all videos, demonstrating the robustness of RoBERTa in capturing sentiments within diverse content. The only performance of MNB that competes with RoBERTa is observed in the video titled "GPT-4 is OVERHYPED," marking it as the highest achievement for MNB with accuracies identical to RoBERTa at 93.2%. Both RoBERTa and MNB methods achieve 93% accuracy. After further analysis of the datasets, it's probably due to clean data with the least noise, and comments are typically written in shorter text compared to those in other datasets. To draw a parallel with the " Two Sum - Leetcode 1 - HashMap - Python " video, it's noteworthy that the comments in that context frequently include pseudo codes, potentially resulting in suboptimal performance for RoBERTa with only 82.9%.

The RoBERTa model achieved its highest accuracy in the video titled "Why Do We Itch," registering an impressive 93.77%, surpassing the Multinomial Naive Bayes (MNB) model with 87.51% accuracy. However, upon closer examination in Table I, we observe that RoBERTa exhibits its lowest accuracy in the video titled "Two Sum - Leetcode 1 - HashMap - Python" from the LeetCode channel with only 82.9%. This particular video, with the smallest dataset among the analyzed videos, highlights the importance of extensive training data for optimal RoBERTa performance.

The observed decrease in accuracy in the mentioned videos suggests that RoBERTa benefits significantly from larger datasets to optimize its predictive abilities. Notably, videos such as "Speed Learning: Learn In Half The Time | Jim Kwik", "My Brain after 569 Leetcode Problems", and "Two Sum - Leetcode 1 - HashMap - Python" share a common trait of having datasets containing fewer than 1000 instances. These instances represent the only scenarios where RoBERTa's accuracy falls below 90%. This consistent trend underscores the close relationship between the model's efficacy and the volume as well as the diversity of training data, emphasizing the necessity of extensive and varied datasets to fully leverage RoBERTa's potential in sentiment analysis for YouTube videos.

However, there are exceptions to this trend. For instance, the video dataset titled "GPT-4 is OVERHYPED," despite having fewer than 1000 instances, not only achieves the second-highest accuracy for RoBERTa but also outperforms all others, claiming the top accuracy spot for MNB as well. Both RoBERTa and MNB methods achieve an accuracy of 93%. Further examination of the datasets suggests that this might be attributed to clean data with minimal noise, where comments tend to be shorter compared to those in other datasets. Drawing a comparison with the "Two Sum - Leetcode 1 - HashMap - Python" video, it's worth noting that comments in that context often contain pseudo-codes, potentially leading to suboptimal performance for RoBERTa with an accuracy of only 82.9%, as demonstrated in Table II.

Table II Evaluation Result on RoBERTa and MNB

Video	Channel	RoBERTa			MNB		
		Accuracy	**F1 Score**	**Precision**	**Accuracy**	**F1 Score**	**Precision**
Why You Are Lonely and How to Make Friends	Kurzgesagt	91.34%	91.34%	91.34%	76.50%	67.00%	75.00%
The Most Horrible Parasite: Brain Eating Amoeba	Kurzgesagt	92.68%	92.54%	92.52%	79.39%	71.00%	83.00%
What Actually Happens When You Are Sick?	Kurzgesagt	91.75%	91.54%	91.40%	85.42%	79.00%	73.00%
Speed Learning: Learn In Half The Time | Jim Kwik	Mindvalley	87.04%	86.66%	86.37%	75.71%	67.00%	71.00%
Unleash Your Super Brain To Learn Faster	Mindvalley	92.14%	91.95%	91.88%	80.43%	72.00%	82.00%
Change Your Frequency, Change Your Reality	Mindvalley	90.14%	89.79%	89.77%	76.14%	70.00%	79.00%
The power of vulnerability | Brené Brown	Ted	90.14%	90.03%	89.96%	77.61%	68.00%	79.00%
How not to take things personally? | Frederik Imbo | TEDxMechelen	Ted	92.58%	92.66%	92.76%	82.10%	78.00%	82.00%
How a 13 year old changed 'Impossible' to 'I'm Possible' | Sparsh Shah | TEDxGateway	Ted	90.11%	90.16%	90.25%	78.41%	75.00%	78.00%
How An Igloo Keeps You Warm	Be Smart	91.96%	91.93%	91.94%	84.62%	78.00%	72.00%
What Is Fire?	Be Smart	90.77%	90.67%	90.66%	71.28%	60.00%	80.00%
Why Do We Itch?	Be Smart	93.94%	93.80%	93.77%	87.51%	82.00%	85.00%
My Brain after 569 Leetcode Problems	NeetCode	87,9%	86,9%	85,8%	82,5%	83,0%	83,0%
Two Sum - Leetcode 1 - HashMap - Python	NeetCode	82,9%	80,6%	81,5%	78,7%	73,0%	81,0%
GPT-4 is OVERHYPED	NeetCode	93,2%	89,9%	86,9%	93,2%	90,0%	87,0%

V. CONCLUSION AND FUTURE WORK

Our study explored sentiment analysis in educational YouTube videos using the RoBERTa model and compared it to Multinomial Naive Bayes (MNB). RoBERTa showcased its effectiveness, achieving its average accuracy at 91.21%, surpassing MNB at 79.59%.

RoBERTa's accuracy dropped below 90% in videos with datasets below 1000, such as "Speed Learning: Learn In Half The Time | Jim Kwik," "My Brain after 569 Leetcode Problems," and "Two Sum - Leetcode 1 - HashMap - Python." This highlights the importance of large, diverse, and quality training data for RoBERTa's optimal performance. In summary, our findings highlight the importance of the dataset's characteristics and size to build a robust model using RoBERTa. The study contributes insights to sentiment analysis in educational multimedia, guiding future research in leveraging sophisticated natural language processing models for sentiment understanding in diverse online content.

Future work for This research involves expanding the dataset by introducing additional classes. This expansion aims to empower the model not only to categorize comments as neutral, positive, or negative but also to identify the subject of the comment, distinguishing whether it refers to the video's content or the topic being discussed. Additionally, the goal is to enhance the model's capability to predict sarcasm and irony, providing a more comprehensive understanding of the nuanced sentiments expressed in educational YouTube videos.

REFERENCES

[1] Osman, W., Mohamed, F., Elhassan, M. et al., "Is YouTube a reliable source of health-related information? A systematic review," *BMC Med Educ,* 2022.

[2] Halim, Z., Hussain, S., Hashim Ali, R., "Identifying Content Unaware Features Influencing Popularity Of Videos On Youtube: A Study Based On Seven Regions," *Expert Syst. Appl.,* p. 206, 2022.

[3] Moghavvemi, S., Sulaiman, A., Jaafar, N. I., & Kas, "Social Media As A Complementary Learning Tool For Teaching And Learning: The Case Of Youtube," *International Journal of Management in Education,* pp. 37 - 42, 2018.

[4] Gu, C. Lin, S. Sun, J. Yang, C. Chen, J. Jian, "What Do Users Care About? Research On User Behavior Of Mobile Interactive Video Advertising," *Heliyon,* p. 8, 2022.

[5] Medhat, W., Hassan, A., Korashy, H.,, "Sentiment Analysis Algorithms And Applications: A Survey," *Ain Shams Engingeering Journal*, pp. 1093 - 1113, 2014.

[6] A. Sadia, F. Khan, and F. Bashir,, "An Overview Of Lexicon - Based Approach For Sentiment Analysis," *International Electronic Engineering Conference*, pp. 1 - 6, 2018.

[7] B. I. Permata, A. E. Prihatini, and Widiartanto, "Pengaruh Kualitas Produk Dan Periklanan Youtube Terhadap Loyalitas Pengguna Brand Wardah Kosmetik Di Kota Semarang," *Jurnal Ilmu Administrasi Bisnis*, vol. 7, pp. 250 - 257, 2018.

[8] N. Riyastika, "Analisis Penggunaan Youtube Pemerintah ProvinsiDKI Jakarta Sebagai Sarana Komunikasi Ditinjau Dari Sudut Pandang Political PR," *Library UI*, p. 21, 2014.

[9] S. Christina and D. Ronaldo, "A Survey of Sentiment Analysis," *A Survey ot Sentiment Analysis Using Sentirwordnet Bhs. Indonesia*, vol. 12, pp. 201 - 241, 2018.

[10] D. Ajeng and L. Marlinda, "Comparison of SVM & Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter," *International Conference Cyber IT Service Manager*, pp. 1 - 6, 2018.

[11] Tanesab, F. I., Sembiring, I., Purnomo, H. D., "Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine," *International Journal of Computer Science and Software Engineering (IJCSSE)*, vol. 6, 2017.

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

[13] Samsir, Kusmanto, A. H. Dalimunthe, R. Aditiya and R. Watrianthos, "Implementation Naïve Bayes Classification for Sentiment Analysis on Internet Movie Database," *Building of Informatics, Technology and Science (BITS)*, vol. 4, pp. 1-6, 2022.

[14] X. Wang, S. Chen, T. Li, W. Li, Y. Zhou, and J. Zh, "Depression Risk Prediction for Chinese Microblogs via Deep-Learning Methods : Content Analysis," *National Library of Medicine*, pp. 1 - 10, 2020.

[15] Xu, F., Pan, Z., & Xia, R, "E-Commerce Product Review Sentiment Classification Based On A Naïve Bayes Continuous Learning Framework," *Information Processing & Management*, 2020.

[16] Alkaff, M., Rizky Baskara, A., & Hendro Wicaksono,, "Sentiment Analysis of Indonesian Movie Trailer on YouTube Using Delta TF-IDF and SVM," *5th International Conference on Informatics and Computing*, 2020.

[17] Mathur, A., & Foody, G. M., "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geoscience and Remote Sensing Letters*, pp. 241 - 245, 2008.

[18] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 2019.

[19] J. Fernando and Fathoni, "Analisis Sentimen Pada E-Learning Universitas Xyz Menggunakan Metode Naïve Bayes Classifier dan Support Vector Machine," *Jurnal Ilmiah Teknologi Informasi (JUTI)*, vol. 21, pp. 81 - 97, 2023.

[20] Y. Liu et al, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," 2019.

[21] B. Richardson, "Implementasi Indobert-Lite Dan Roberta Untuk Text Mining Pada Aplikasi Chatbot Jacob," 2021.

[22] Samsir, Ambiyar, Verawardina and Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *J. Media Inform. Budidarma*, vol. 5, pp. 157-163, 2021.

[23] Samsir et al., "Naives Bayes Algorithm for Twitter Sentiment Analysis," *J. Phys. Conf. Ser.*, 2021.

[24] S. Wang and L. Jiang, "Adapting Naive Bayes Tree for Text Classification," *Knowledge and Information Systems*, 2015.