Unveiling Sentiments: Analyzing Learner's experience using VADER and RoBERTa models

1st Navaneetan M
Computer Science and Engineering
Hindustan Institute of Technology & Science
Chennai, India.
navaneetan973@gmail.com

2nd Tharagesh G
Computer Science and Engineering
Hindustan Institute of Technology & Science
Chennai, India.
tharagesht@gmail.com

3rd A Sai Sabitha
Computer Science and Engineering
Hindustan Institute of Technology & Science
Chennai, India.
asais@hindustanuniv.ac.in

2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) | 979-8-3503-6482-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ADICS58448.2024.10533527

***Abstract*—Educational Institutions receive a large amount of data, which includes student review. The text-based interactions, reveal patterns in student sentiments and emotions and enhance the overall educational experience. The text analytics allows institutions to perform analysis on these feedbacks, that gives insights to the institution on student satisfaction, identifying areas of improvement, and analyzing the students' response. The Sentiment analysis is performed on reviews to help institutions understand the student needs and their engagement in courses in a MOOC environment. The sentiment analysis models are used to classify positive or negative reviews or to determine emotions such as happy or sad. In this research work two approaches, lexicon and transformer based algorithms were used to understand the review of Coursera data, The analysis further shows sentiment of learners based on different domains of study.**

Index Terms*—VADER (Valence Aware Dictionary and Sentiment Reasoner), Massive Open Online Courses, Sentiment Analysis, RoBERTa (Robustly optimized BERT approach), BERT (Bidirectional Encoder Representations from Transformers)

I. Introduction

The Massive Open Online Courses (MOOCs) have been in trend and interest of researchers for quite a few years. Their has been helpful in choosing the right courses for the stu- dents and professionals [1]. The Students have been choosing MOOC courses based on the influence of participation and their interaction on a particular course and also situational interest plays a major role in intrinsic motivation and partici- pation on performance [2]. The Students prefer online courses based on several aspects which include the reputation of the course provider, the faculty handling the course, interactivity in the course and mainly based on the user feedback. In the con- text of MOOC courses user feedback is also known as reviews, they play a major role in the selection of courses. The recent trends in MOOCs have gained momentum, As mentioned,the focus revolves around the extraction and analysis of reviews provided by learners who have participated in various Coursera courses. The reviews in this research work were extracted from several courses provided by learners, specifically the domain with higher number of courses has been selected and categorized into technical and medical courses and the text mining has been performed using their reviews. The study unfolds into three distinct case studies, each contributing to a comprehensive understanding of sentiment analysis algorithms and their practical implications in the context of course review analysis:

Exploration of Sentiment Analysis Algorithms: The first case study delves into the landscape of sentiment analysis algo- rithms. The evaluation and comparison of several algorithms has been performed to discern their strengths and weaknesses in gauging sentiments expressed in course reviews.

Temporal Analysis of 2019 and 2020 Course Reviews: The second case study, focuses on the sentiment dynamics within the course reviews from the years 2019 and 2020. Insights were gained in evolving learner perceptions and experiences, by examining how the sentiments have evolved over time.

Comparative Analysis across Categories: The third case study takes a granular approach by categorizing reviews into distinct domains, such as Technical and Medical. By comparing the sentiment patterns across these categories, the aim is to un- cover the nuances in learner sentiments specific to different course domains.

II. Related Works

The recent scholarly work has given significant emphasis to the application of text mining algorithms in the study of user reviews within the context of Massive Open Online Courses (MOOCs). The study by [3] used text mining to analyze online reviews of business courses on Coursera to shed light on the variables impacting the enhancement of the learning experience.

The COVID-19 period has significantly improved the funding, budget and opportunity to learn in MOOC's [4]. The learners have shown more interest during and after the period and significant amount of budget has been allocated in Bangladesh, The Bangladesh government had proposed to allocate USD 8.3 billion of the budget to the education sector for the fiscal year 2021-22, which was 11.9 % of the total budget and is only 2.09% of the GDP, The market was valued between USD 700–800 million and is expected to grow up to USD 30 billion in the next decade. In fact, MOOCs have been more convenient, cost effective and promote pluralism and conceptual understanding of diverse resources [4]. Addition- ally, [5] who discussed for a recommendation system based on student comments, demonstrating the usefulness of text mining in enhancing the course recommendations. Together, these findings highlight the expanding importance of text mining techniques in the examination of user reviews on Coursera, which provides an insightful information for improving the online education.

A. Lexicon based algorithms [6]

Lexicon based sentiment analysis models rely on predefined lexicons or dictionaries that contain words or phrases that are related to 3 predominant sentiment scores, mainly positive, negative and neutral. These lexicons and their corresponding scores determine their sentiment polarity. That is, the words that convey positive sentiment receive positive scores, those conveying negative sentiment receive negative scores, and the neutral words are assigned the score zero. The Lexicon-based sentiment analysis [7] is suitable for scenarios where a quick, interpretable analysis of sentiment are needed, especially in cases of informal or social media text and social reviews . [8] Throughout the research work, the authors aim to adopt rule-based approaches that define a set of rules and utilize inputs such as classic Natural Language Processing techniques, stemming, tokenization, a region of speech tagging, and machine learning parsing for sentiment analysis.

[9] The researchers proposed a generic, extendable, domain-independent lexicon-based framework for sentiment analysis that overcome scalability issues, limitations and outperform the existing models in terms of accuracy. The framework's engine aggregates sentiment intensities using various lexicons, and a dedicated neutral bias module manages misinterpreted neutral polarities. The extensive experiments conducted on five different lexicons and utilize a sample sentiment engine demonstrated significant improvements in accuracy compared to traditional lexicon-based sentiment approaches.

1) VADER [10]: The Valence aware dictionary for sentiment reasoning (VADER) lexicon performs exceptionally well in the social media domain. The correlation coefficient shows that VADER (r = 0.881) performs, as well as individual human raters (r = 0.888) at matching ground truth (aggregated group mean from 20 human raters for sentiment intensity of each tweet). Vader was introduced by [11], the algorithm scales sentiments from –4 to +4 scale where –4 is extremely negative sentiment, +4 is extremely positive sentiment and 0 is neutral sentiment. [11]They conducted a comparative analysis between the VADER lexicon and seven other well known sentiment analysis lexicons. To ensure fairness in the comparisons, all assessments employed VADER's rule-based model for processing syntactic and grammatical cues. The only distinction among the comparisons lay in the features represented within the respective lexicons. The findings indicate that VADER's lexicon demonstrates exceptional performance in the realm of social media and exhibits favourable generalization.

2) SentiWordNet [12]: The SentiWordNet is a lexical resource that extends WordNet (a lexical database of English) by assigning sentiment scores to words. It provides numerical scores representing the positivity, negativity, and neutrality of words in the English language. The Sentiwordnet was introduced by [13]. The reliance of SentiWordNet scores on term glosses emerges as a potential constraint in the precision of term scores and the overall classification accuracy of this method [14]. In the final analysis, factors such as the utilization of colloquial language and expressions in the absence of opinion information, the disambiguation of WordNet terms with multiple meanings, inaccuracies in the assignment of part-of-speech tags, and the accurate identification of named entities like actor and film names have been recognized as contributors to misclassifications observed through the utilization of this method.

B. Transformer based algorithm [15]

1) RoBERTa: Robustly Optimized BERT approach, was introduced by [16], The author found that BERT [17] was undertrained. They carefully evaluated the effects of hyperparameter tuning & observed that performance improved when training the model with longer sequence & bigger batches. They used a novel dataset, CCNEWS, more data for pre-training, further improved the performance on down stream tasks. A study by [18] proposed a multi task aspect categorising sentiment analysis model based on RoBERTa. The deep bidirectional transformer that can extract from both aspect tokens and features. A unique study by [19] combined RoBERTa and BiGRU (Bidirectional Gated Recurrent Unit) for large corpus of Vietnamese text for sentiment analysis. RoBERTa has been used for clothing and fashion, A study by [20], used calibrated RoBERTa for multilabel classification, BERTopic for topic modelling and this method had proven results in sustainability and waste reduction management in the fashion industry.

III. METHODOLOGY

The Fig 1 represents the overall architecture used in this study.

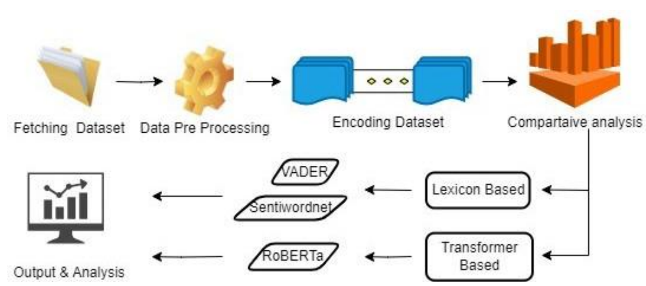

Fig. 1. Proposed architecture

1. Collection of dataset from Kaggle website (https://www.kaggle.com/datasets/imuhammad/course-reviews-on-coursera) through google dataset search (https://datasetsearch.research.google.com/)

2. Preprocessing the data through data cleaning and preprocessing methods.

3. Sentimental analysis with the cleansed dataset.

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE

4. Lexicon based Analysis under dictionary-based sentiment analysis. That is VADER (Valence Aware Dictionary and sentiment Reasoner) analysis.
5. Transformer based Analysis using Robustly optimized Bidirectional Encoder Representations from Transformers(RoBERTa).
6. Extraction of sentiment score. Which are positive, negative and neutral sentiments.

IV. Experimental setup

A. Data set

The Google dataset search was used for the case study for finding dataset. For the use of text mining, we wanted reviews of courses, for which we used the dataset which had Coursera course reviews in Kaggle. Coursera is one of the largest Massive Open Online Courses (MOOC) providers in the world and it hosts numerous courses from top universities across the globe and in different subjects. The dataset providers collected over 1.45 million course reviews posted by students and participants.

	A	B	C	D	E	F	G
1	reviews	reviewers	date_revie	rating	course_id		
2	Pretty dry, but I was able to pass with ju	By Robert S	12-Feb-20	4	google-cbrs-cpi-training		
3	would be a better experience if the vide	By Gabriel E	28-Sep-20	4	google-cbrs-cpi-training		
4	Information was perfect! The program i	By Jacob D	08-Apr-20	4	google-cbrs-cpi-training		
5	A few grammatical mistakes on test ma	By Dale B	24-Feb-20	4	google-cbrs-cpi-training		
6	Excellent course and the training provic	By Sean G	18-Jun-20	4	google-cbrs-cpi-training		

Fig. 2. Coursera reviews data

	A	B	C	D	E	F
1	name	institution	course_url	course_id		
2	Machine Learning	Stanford University	https://www.cou	machine-learning		
3	Indigenous Canada	University of Alberta	https://www.cou	indigenous-canada		
4	The Science of Well-Being	Yale University	https://www.cou	the-science-of-well-being		
5	Technical Support Fundament	Google	https://www.cou	technical-support-fundan		

Fig. 3. Coursera courses id

The kaggle repository Table 1 had two datasets, where Fig 2 named "Coursera reviews data" had attributes like "Reviews, Reviewers, Date_reviews, Rating, Course_ID" Similarly Fig 3 named "Coursera courses data" had attributes like "Name, Institution, Course_url,Course_ID"

TABLE 1
Meta data of dataset

Name of meta data	Coursera.csv
Meta Data Platform	Coursera
Total No.of Columns	5
Total No.of Rows	10,00,000+
No.of attributes	4
Categories	Medical, Technical

B. Data preprocessing and cleaning

In the early stages of this investigation, focus was placed on the important process of data preparation and cleansing. The dataset's integrity and quality are essential to the accuracy and dependability of any subsequent text analysis. To achieve this, the raw dataset underwent several significant rounds of refinement.

1) Removal of Noisy data: The process of data preprocessing also entailed the removal of irrelevant data entries. These entries included records that did not pertain to the research objectives or were considered outliers. In the data set we had also removed the irrelevant data before 2019 in the "date‾reviews" column, that is only the 2019 and 2020 data were considered for the study and research. By systematically identifying and removing such data, the dataset's focus was sharpened, and the analysis was better poised to extract meaningful insights.

2) Handling Empty Columns: Another essential aspect of data preprocessing was the identification and management of empty or missing columns. A systematic process of detecting and handling empty columns was executed, where columns with a significant proportion of missing values were either imputed or excluded, depending on the nature of the analysis. This meticulous approach ensured that only complete and meaningful data were retained for further investigation.

3) Selection of Attributes: In the raw dataset, a multitude of columns were present, containing various attributes and metadata. As part of the preprocessing phase, a careful evaluation of these columns identified those that held no direct relevance to the research objectives. The Columns such as "reviewers" contained PII(Personally Identifiable Information) and were subsequently removed. The elimination of irrelevant columns not only streamlined the dataset but also reduced the dimensionality, thereby enhancing the efficiency of subsequent analysis.

C. Domain categorization

Based on the course name, this study has separated the dataset manually as Technical and Medical Domains, with the majority of courses in the dataset falling these categories.

D. Rapid miner and colab tool

In this Analysis, a comprehensive experimental setup was employed to analyze the dataset and derive valuable insights from it. The process involved two primary tools: Colab and RapidMiner. The Fig 4 RapidMiner was utilized for initial data preprocessing and basic exploratory data analysis (EDA). The software enabled data cleaning, sorting, and visualization, facilitating an initial understanding of the dataset's characteristics. Subsequently, Fig 5 Colab, a powerful data mining and machine learning platform, was employed for more advanced data analysis tasks. The combination of these tools ensured a robust and systematic approach to the case study, enabling us to uncover hidden patterns and valuable insights within the text data.

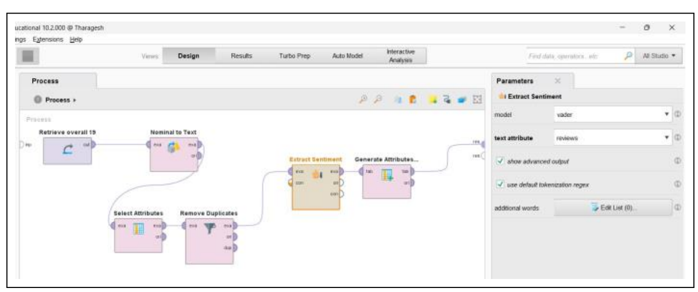

Fig. 4. RapidMiner workspace

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE

V. Result and Analysis

A. Case Study 1

Comparative study of Sentiment analysis algorithms: The preprocessed dataset was considered for analysis of three algorithms namely Sentiwordnet, VADER and RoBERTa.

1. The analysis of lexicon based algorithms (VADER & Sentiwordnet) were performed in Colab. The results are shown below in Table 2.

2. The analysis of Transformer based algorithm (RoBERTa) were similarly performed in Colab. The results are shown below in Fig 5.

```
# Calculate accuracy and precision
accuracy = (positive_count + negative_count) / len(results_df)
precision = positive_count / (positive_count + negative_count)

# Print the results
print("Accuracy:", accuracy)
print("Precision:", precision)

Accuracy: 0.8082236550250911
Precision: 0.9135992227701365
```

Fig. 5. Output of RoBERTa

TABLE 2
Performance metrics of algorithms

Metrics	Sentiwordnet	VADER	RoBERTa
Accuracy	47%	75%	80%
Precision	43%	79%	91%
F1 Score	45%	82%	85%

From the above comparison, RoBERTa outperformed the other two algorithms in metrics like accuracy, precision and F1-Score, however, RoBERTa took a huge amount of time to train the data(3 hours) in Google Colab using T4 GPU and only had two sentiment classifications which were positive and negative sentiments. Accuracy was calculated based on these two sentiments, whereas VADER gave positive, negative and neutral sentiments.

B. Case study 2

Comparison of year 2019 and year 2020 sentiment analysis: It is clear from Case study 1 that RoBERTa has outperformed VADER in the means of accuracy and precision, however VADER has higher sentiment nuances nearly 9 sentiment classifications. Hence this case study uses VADER for better understaning of learner's sentiments.

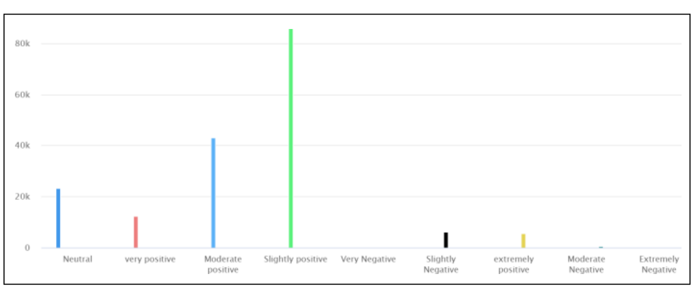

Fig. 6. Sentiment analysis of the year 2019

Fig 6 represents the sentimental analysis of the Coursera reviews of the year 2019, which has 6.95% of very positive, 3.56% of slightly negative and 3.17% of extremely positive sentiments. The Fig 7 represents the sentimental analysis of

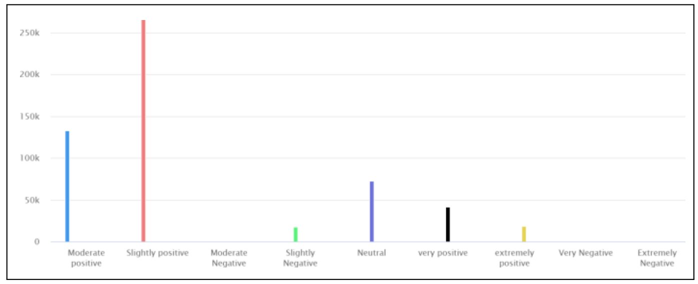

Fig. 7. Sentiment analysis of the year 2020

the Coursera reviews of the year 2020,which has 7.50% of very positive, 3.25% of slightly negative and 3.44% of extremely positive sentiments. The attributes were classified based on the score which was assigned by the VADER to the string and tokens ranging from +4 to -4, slightly positive = +1, moderate positive = +2, very positive = +3, extremely positive =+4, neutral = 0, slightly negative = -1, moderate negative = -2, very negative = -3, extremely negative = -4. Based on the score, the tokens were classified according to their respective attributes. From Table 3, the data of sentiment analysis of the year 2019 and sentiment analysis of the year 2020. The case study infers that there was a significant increase of 0.55% in very positive sentiments, 0.27% increase in extremely positive sentiments and 0.31% decrease in slightly negative sentiments from the previous year.

TABLE 3
Sentiment analysis using VADER

Sentiments	2019	2020
slightly positive	48.47%	48.27%
Very positive	6.95%	7.50%
moderately positive	24.34%	24.14%
extremely positive	3.17%	3.44%
neutral	13.17%	13.14%
slightly negative	3.56%	3.25%
Very negative	0.041%	0.024%
moderately negative	0.26%	0.19%
extremely negative	0.010%	0.005%

C. Case study 3

A comparative sentimental analysis of data based on categorical manner such as technical and medical. The Fig 8

```
positive_count = df['roberta_pos'].sum()
negative_count = df['roberta_neg'].sum()
neutral_count = df['roberta_neu'].sum()

print("Total positive sentiment:", positive_count)
print("Total negative sentiment:", negative_count)
print("Total neutral sentiment:", neutral_count)

Total positive sentiment: 46052.39036035631
Total negative sentiment: 4686.912038062175
Total neutral sentiment: 12116.697598112281
```

Fig. 8. Medical Domain

represents the sentimental analysis of the reviews which is categorized into technical, which gives three attributes namely positive, negative and neutral. In the analysis, the positive sentiment count is 46052, negative sentiment is 4686 and neutral sentiment is 12116. The Fig 9 represents the sentimental

```
positive_count = results_df['roberta_pos'].sum()
negative_count = results_df['roberta_neg'].sum()
neutral_count = results_df['roberta_neu'].sum()

print("Total positive sentiment:", positive_count)
print("Total negative sentiment:", negative_count)
print("Total neutral sentiment:", neutral_count)

Total positive sentiment: 60306.83746042603
Total negative sentiment: 3440.284909226466
Total neutral sentiment: 10853.87757443171
```

Fig. 9. Technical Domain

analysis of reviews categorized into technical, which gives three attributes namely positive, negative and neutral. In this research work, the positive sentiment count is 60306, negative sentiment is 3440 and neutral sentiment is 10853. This case study infers that the Technical dataset demonstrates a slightly greater prevalence of highly positive reviews when contrasted with the medical dataset.

VI. Conclusion and future work

In conclusion, the research work has provided a comprehensive exploration into the lexicon and transformer based facet of sentiment analysis. Through three distinct case studies, we have illuminated various dimensions of sentiment analysis. The first case study involved a comparative and algorith- mic analysis of three algorithms Sentiwordnet, VADER and RoBERTa, revealing valuable insights into their individual strengths and limitations. In the second case study, the applica- tion of VADER to analyze reviews from 2019 to 2020 unveiled discernible shifts in sentiment trends over time. Lastly, the third case study categorizes courses into Medical, and Tech- nical domains, offering nuanced insights into the sentiments associated with each.The investigations have demonstrated that the Technical dataset exhibits a marginally higher prevalence of very positive reviews compared to the medical dataset, indicating a noteworthy trend in sentiment distribution across diverse course categories. Overall, these findings contribute to a deeper understanding of sentiment dynamics in online platforms, emphasizing the importance of transformer and lexicon-based approaches in uncovering nuanced sentiments across varied contexts. GPT and XLNET based analysis should be a focus of future studies. Analyzing the cause and influence of trends in sentiments in a particular year can give insights into learner's sentiments.

References

[1] Gupta, S., Sabitha, A. S., & Chowdhary, S. K. (2021). Predicting the efficiency and success rate of programming courses in MOOC using machine learning approach for future employment in the IT industry. Journal of Information Technology Research (JITR), 14(2), 30-47.

[2] de Barba, P. G., Kennedy, G. E., & Ainley, M. D. (2016). The role of students' motivation and participation in predicting performance in a MOOC. Journal of Computer Assisted Learning, 32(3), 218-231.

[3] Wei, X., & Taecharungroj, V. (2022). How to improve learning experience in MOOCs an analysis of online reviews of business courses on Coursera. The International Journal of Management Education, 20(3)

[4] Amit, S., Karim, R., & Kafy, A. A. (2022). Mapping emerging massive open online course (MOOC) markets before and after COVID-19: A comparative perspective from Bangladesh and India. Spatial Information Research, 30(5), 655-663.

[5] Hazar, M. J., Zrigui, M., & Maraoui, M. (2022). Learner comments-based Recommendation system. Procedia Computer Science, 207, 2000-2012.

[6] Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".

[7] Lei, L., Liu, D. (2021). Conducting Sentiment Analysis. (n.p.): Cambridge University Press.

[8] Mitra, A. (2020). Sentiment analysis using machine learning approaches (Lexicon based on movie review dataset). Journal of Ubiquitous Computing and Communication Technologies (UCCT), 2(03), 145-152.

[9] Moussa, M. E., Mohamed, E. H., & Haggag, M. H. (2020). A generic lexicon-based framework for sentiment analysis. International Journal of Computers and Applications, 42(5), 463-473

[10] Keikhosrokiani, P., & Pourya Asl, M. (Eds.). (2022). Handbook of research on opinion mining and text analytics on literary works and social media. IGI Global.

[11] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).

[12] Murshed, M. (2013). Technical Challenges and Design Issues in Bangla Language Processing. United States: IGI Global.

[13] Sebastiani, F., & Esuli, A. (2006, May). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of the 5th international conference on language resources and evaluation (pp. 417-422). European Language Resources Association (ELRA) Genoa, Italy.

[14] Ohana, B., & Tierney, B. (2009). Sentiment classification of reviews using SentiWordNet.

[15] Tunstall, L., Werra, L. v., Wolf, T. (2022). Natural Language Processing with Transformers, Revised Edition. United States: O'Reilly Media.

[16] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[17] Ravichandiran, S. (2021). Getting Started with Google BERT: Build and Train State-of-the-art Natural Language Processing Models Using BERT. United Kingdom: Packt Publishing.

[18] Liao, W., Zeng, B., Yin, X., & Wei, P. (2021). An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. Applied Intelligence, 51, 3522-3533.

[19] Trang, N. T. M. (2023). A Hybrid Model of RoBERTa and Bidirectional GRU for Enhanced Sentiment Analysis. International Journal of Open Information Technologies, 11(11), 19-24.

[20] Alamsyah, A., & Girawan, N. D. (2023). Improving Clothing Product Quality and Reducing Waste Based on Consumer Review Using RoBERTa and BERTopic Language Model. Big Data and Cognitive Computing, 7(4), 168.

IEEE 979-8-3503-6482-8/24/$31.00 ©2024 IEEE