ASPECT-BASED SENTIMENT ANALYSIS OF OPEN-ENDED RESPONSES IN STUDENT COURSE EVALUATION SURVEYS

by

Reem ALOTAIBI *

Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Original scientific paper
https://doi.org/10.2298/TSCI2406037A

Student surveys are tools used to gather feedback and opinions from students regarding various aspects of their educational experience such as satisfaction with courses, instructors, supervision, facilities, services, and overall campus environment. These surveys play a crucial role to assess and improve the quality of education and overall student experience, and provide insights that can be used to decision makers. There are two primary categories of survey questions: open-ended and closed-ended. Open-ended questions allow respondents to provide detailed answers in their own words, while closed-ended questions offer only predetermined options. For analyzing student surveys, it is common practice to focus on closed-ended questions, while often open-ended responses provide valuable insights. In this paper, we use sentiment analysis to extract students' emotions towards various aspects of their educational experience using transformer-based pre-trained language models. The results show that the multilingual XLM-RoBERTa demonstrated encouraging results compared to the multilingual BERT model with an accuracy of 0.81 and F1 of 0.80 for the classification of aspects while for the classification of sentiments, obtained 0.94 and 0.92 for accuracy and F1, respectively.

Key words: *student surveys, topic modeling, sentiment analysis, BERT, course evaluation, open-ended responses, XLM-RoBERTa, aspect-based sentiment analysis*

Introduction

Student surveys are a vital tool for evaluating and improving courses in higher education. These surveys provide direct feedback from students on their learning experiences, course content, teaching methods, and overall satisfaction. By gathering and evaluating this feedback, educational institutions can identify strengths and areas for improvement in their academic programs.

Survey analysis in educational contexts typically focuses on close-ended questions due to their straightforward nature and ease of quantification. Close-ended questions provide respondents with a set range of answers, such as multiple choice, scales, or yes / no options. This format allows for efficient data collection and analysis, facilitating the identification of trends, patterns, and statistically significant insights. Structured responses make it easy to aggregate data, perform statistical analyzes, and generate reports that can inform decision-making processes [1, 2].

* Author's e-mail: ralotibi@kau.edu.sa

However, open-ended questions, which invite respondents to express their thoughts in their own words, have significant potential to uncover valuable, detailed information that close-ended questions might miss. Open-ended questions allow students to provide in-depth comments, elaborate on their experiences, and suggest specific improvements. These qualitative data can reveal the underlying reasons behind the trends observed in the quantitative data, providing a deeper understanding of the perceptions and experiences of the students [3, 4].

Aspect-based sentiment analysis (ABSA) has two primary tasks, aspect extraction and sentiment classification [5]. Aspect extraction involves identifying and extracting specific attributes or features known as aspects of an entity from texts. Sentiment analysis identifies and analyzes sentiments represented in textual data [6]. The ABSA has become a hot topic for research in recent years and has attracted a lot of attention, including social media posts, product and customer reviews. Furthermore, sentiment analysis is helpful in the educational field, as it can identify positive, negative, or neutral emotions on instructor performance, learning facilities, and course evaluation [7].

In this research, the main objective of ABSA is to extract aspects, uncover opinions, extract the sentiments they convey, and classify these sentiments into polarities [8]. With recent advancements in ABSA research, there has been a shift towards improving sentiment polarity detection and classification. Additionally, the context of natural language processing (NLP) has seen significant developments, particularly with the experimentation with deep learning (DL) models. Consequently, the emergence of ABSA is a logical progression in this area [9].

This study aims to explore the transfer learning approach and investigate ABSA's performance in the Arabic educational context. The multilingual base model BERT (cased) [10] and XLM-RoBERTa [11], which is an extension of RoBERTa as an improvement over BERT, have been used for the extraction of aspect from open-ended course evaluation surveys. The sentiment of each aspect was then classified using the sentiment classifier and tested by comparing its predictions against the expert-annotated sentiment labels. To our knowledge, this is the first study to utilize ABSA for course evaluation survey responses in Arabic.

Based on previously discussed issues, this paper aims to:

- Apply unsupervised learning approach using BERTopic for topics and aspect extraction.
- Implement an aspect classifier to assign topic names to unseen survey responses. The performance of the classifier is evaluated using the test set.
- Implement a sentiment classifier to predict sentiment labels for unseen survey responses.
- Train a suggestion classifier that detects whether an aspect mentioned in a sentence is missing, needs enhancements, or needs to be removed.

Related work

Within the field of education, ABSA has become a highly effective tool to obtain a deep understanding of a variety of aspects of the learning experience, ranging from course contents and teaching strategies to course instructors and assessment methods. Through detailed analysis of student feedback, ABSA helps decision makers, including administrators and teachers, better understand the viewpoints of their students and implement data-driven decisions.

The analysis of student feedback, course evaluations, and online reviews of the teaching and learning experience are the main sources of sentiment analysis in the context of education. Several approaches, such as machine learning, DL, and hybrid methods, have been used in recent studies. As an example, Sindhu *et al.* [12] used the supervised aspect-based opinion mining method based on the two-layered LSTM model to analyze the teaching skills of teachers. The model achieved 91% accuracy in aspect extraction and 93% accuracy in sentiment classification. Kastrati *et al*. [13], Sultani and Deneshpour [14] has also utilized the LSTM model in addition to weakly supervised annotation to extract aspects and sentiments from student feedback on massive open on-line courses (MOOC). The results show that the model leads to more accurate results than the sentiment analysis methods that depend on manual/human annotation.

In another study, Hussain *et al*. [15] proposed a novel approach for analyzing student feedback on teaching and learning experience, using a multi-layer topic modeling and hybrid machine learning framework named Aspect2Labels (A2L). The method achieved 91.3% accuracy in annotating unlabeled student feedback and over 90% in sentiment classification using ANN and support vector machines (SVM). Moreover, Ren *et al*. [16] adopted DL techniques for aspect-level sentiment analysis to automatically determine the sentiment of text-based students' reviews. The proposed solution achieved a $F1$ value of 79%.

The ABSA using pre-trained language models such as bidirectional encoder representations from transformers (BERT) is capable of efficiently analyzing student responses to locate exactly areas/issues that require development [17]. Zhang *et al*. [18] implemented a BERT-based model for sentiment classification of Chinese online course review MOOC, the results show that the $F1$ reaches 92%. Considering Arabic language and according to a recent study, Alshaukh *et al*. [19] evaluated the efficacy of several BERT-based models for aspect extraction for Arabic text taken from surveys provided by an educational institution. The survey aimed to assess resources, services, and student learning outcomes, the data annotation was performed by domain experts to extract training data aspects. The results demonstrated that the proposed model obtained $F1$ score of 0.58 and 0.86. for the aspect extraction and sentiment classification, respectively.

While the literature on ABSA using BERT is still expanding, applying these techniques specifically for Arabic educational contexts are limited. Most of the research has focused on commercial reviews [20-24], social media texts [25], and news articles rather than educational content. This suggests that there is a need for additional focused study in this field. More sopisticahted model such as XLM-RoBERTa is a powerful transformer-based model that has shown significant promise in various NLP tasks across multiple languages. In the educational context, XLM-RoBERTa can be particularly effective in ABSA within education, where it helps in extracting specific aspects such *curriculum*, *instructors*, o *teaching method*s from student reviews and determining the sentiment expressed towards each aspect.

Materials and methods

This section presents the dataset and models that were used*. Next, we went over the methodology employed in creating the ABSA classifier. Finally, we outlined the performance metrics that were employed in this study to assess the various tasks.

Dataset

The purpose of the course evaluation survey is to evaluate the courses taught in the postgraduate programs of the Faculty of Computing and Information Technology at King Abdulaziz University in the academic years 2020-2024. The questionnaire includes thirty three closed-ended questions categorized into five dimensions and four open-ended questions in Arabic language. The open-ended questions assess various aspects: suggestions to add, what suggestion(s) do you have to improve this course, what did you like most about this course, and what did you dislike most about this course. In this work, we employ four open-ended questions written in Arabic resulting in a final dataset of 3078 responses and four features. For aspect

* https://github.com/ralotaibi/ABSA/

and sentiment classifiers, we use the four columns, while for the suggestion classifier, we only include the two columns related to suggestions. Table 1 shows a few sample answers to the open-ended questions.

Table 1. A sample of students' responses used in the experiment

Suggestions	
Increase the number of hours or reduce the duties to be appropriate	Ease of communicating information and cooperating with students by finding appropriate solutions
Improvement	
Increase the practical part after each chapter	It is better to change the curriculum plan to an organized one instead of distracting students by taking additional courses
Like	
Course instructor is wonderful and very helpful	The curriculum is very useful, I recommend it
Dislike	
Course hours is very long	It was more theoretical than practical

Data pre-processing was performed using AraBERT [26] for Arabic text pre-processing to remove null values, spaces, diacritics and duplicate. In each column, we additionally replace the top 15 frequently occurring words with NaN. The short string which is less than five letters were eliminated for topic modelling. Finally, some sentences include English words and we decided to keep them. After pre-processing step, the final dataset become 1610 records.

Aspect classifier workflow

The first phase is topic modeling, BERTopic is employed to extract topics from the reviews. Similar topics are then merged together. The embedding process utilizes the DistilBERT base multilingual model, which is built on the BERT architecture and produces 512-dimensional vectors. To reduce the dimensionality of these embeddings, Uniform Manifold Approximation and Projection (UMAP) [27] is employed. The reduced embeddings are then clustered using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [28]. Finally, Class-based Term Frequency-Inverse Document Frequency (C-TF-IDF) is used to create topic representations [29], resulting in the identification of 22 distinct topics.

After identifying aspect categories, we utilize the outlier reduction algorithm in BERTopic to handle outliers [30]. We found out that some topics contain reviews belonging to other topics. Therefore, an additional technique is applied, as:

– Embeddings of each review in the dataset are obtained.
– Topic embeddings are calculated by averaging the embeddings of the reviews associated with each topic (excluding the problematic topics).
– Similarity between each review (in the problematic topics and the remaining outliers from the previous step) is computed against all possible topics.
– Topics are assigned based on this similarity calculation.

The final topic list includes nine distinct topics as shown in tab. 2.

Table 2. Topics generated by BERTopic

Topic	Description
Curriculum	Course content
Doctor	Course instructor
Practical	Balance between theoretical and practical contents
Tasks	Course tasks, assignments and grading
Teaching Methods	Teaching methods used in the class
Time Table	Schedule of classes
Labor Market	The course links to the labor market
Research	Research activities in the course
None	–

Data splitting. Once the previous steps have been completed, the data is divided into two separate sets: one for training and another for testing. The training set contains labels assigned by the next step: Automatic Data Annotation. The test set labels are reviewed and overwritten by a subject matter expert. The performance of the aspect classifier will be evaluated against this test set.

Automatic data annotation. We utilized three distinct large language models (LLM): OpenAI's GPT-4o *, Meta's LLama3 70B **, and Cohere's Command R Plus ***. Each example in our dataset was submitted to these LLM, with instructions to assign one label from nine topics identified in previous steps. Consequently, we obtained three labels for each example, one for each LLM. To determine the final label, we applied a majority voting system, with a preference for the GPT-4o label in case of a tie. For the training data, we retained the label derived from the majority vote. For the test data, we forwarded it to the subject-matter expert for review and update as necessary.

Figure 1 shows the distribution of various topics that were extracted from the responses of students to the evaluation of the courses. The *x*-axis lists the different topics, while the y-axis indicates the number of instances or responses associated with each topic.

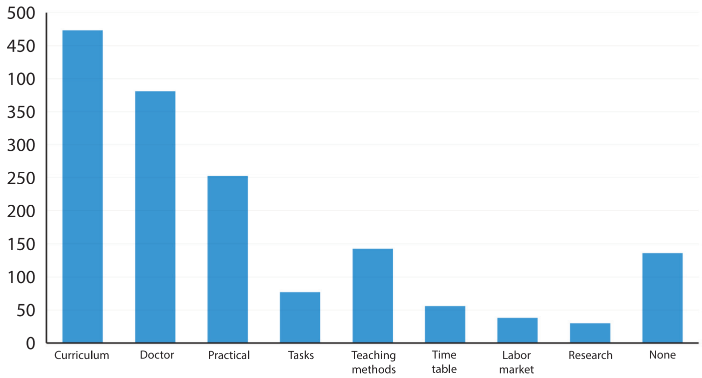


Figure 1. The distribution of various topics extracted from student course evaluation responses

* https://openai.com/index/hello-gpt-4o/

** https://huggingface.co/meta-llama/Meta-Llama-3-70B

*** https://docs.cohere.com/docs/command-r-plus

Data augmentation. We used OpenAI's GPT-4o to generate additional examples from each instance in our training data, ensuring that these new examples are similar to the original while remaining distinct from both the original and each other. We now have a Python dictionary where the keys represent examples from the training data, and the values are the generated sentences corresponding to each example in the training dataset.

To increase the representation of minority classes, we identified examples within each minority class and retrieved their similar instances from the previously generated data. Training data distribution after this step shown in fig. 2.

Aspect classifier training and evaluation. Two aspect classifiers were trained to assign topic labels to unseen reviews after the training data was augmented. BERT multilingual base model (cased)* is the first classifier. The second classifier is XLM-RoBERTa**, a BERT aspect classifier. The test set is used to assess how well these two classifiers perform.

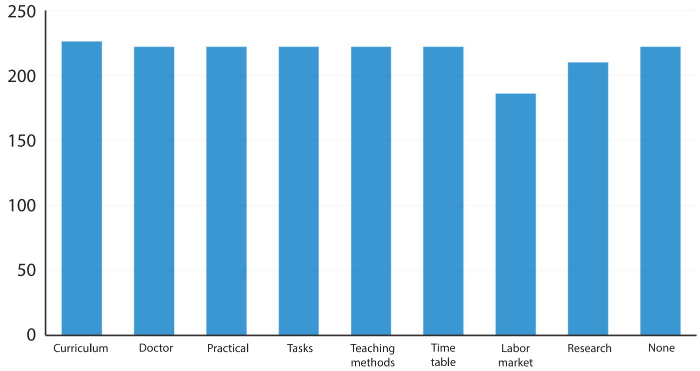


Figure 2. The final topic distribution after data augmentation

Sentiment classifier workflow

For sentiment annotation, we employed the same methodology used in *aspect annotation* to annotate each example in our dataset with one of three labels: Positive, Negative, or Neutral, as seen in fig. 3(a). Once more, the test set was sent to the subject matter expert (after the annotation step of the LLM) for review and any necessary updates. Using the cleansed training data, two sentiment classifiers, BERT and XLM-RoBERTa, are trained to predict sentiment labels for unseen reviews. The performance of the sentiment classifiers is assessed by comparing its predictions against the labels in the test set.

Suggestions classifier workflow

The data are already divided into training and test sets; however, we need to filter it to include only the entries from the suggestion columns in both sets. Again, the same approach is used to annotate the data. The possible labels are: needs to be added, need enhancement, needs to be removed, or none as seen in tab. 3. For augmentation, we used the same method described in the data augmentation step of the topics to augment the training data of suggestions. The distribution of training data after this step is shown in fig. 3(b). We utilized data from the previous step to train the BERT and XLM-RoBERTa classifiers. Performance was evaluated by comparing the predictions with the labels in the test set.

* https://huggingface.co/google-bert/bert-base-multilingual-cased

** https://huggingface.co/FacebookAI/xlm-roberta-large

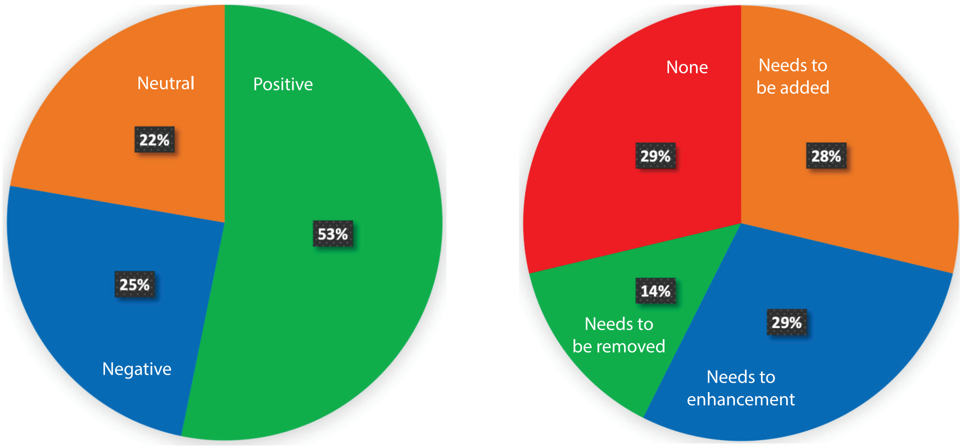


Figure 3. Label distribution; (a) the distribution of sentiment classes based on students' responses and (b) label distribution of students' suggestions after augmentation

Table 3. The distribution of students' suggestions

Label	Description	*n*
Needs to be added	When the student suggests adding something that is currently missing	174
Needs enhancement	Something that exists but in low quality, suggesting improvement	86
Needs to be removed	When the student refers to something existing but suggests its removal	14
None	Otherwise	51

Evaluation metrics

The most widely used metrics in classification are used to assess the proposed models, and these metrics include:

Precision. Precision measures the ratio of topics/classes the model predicted, whether they were correct or incorrect, to the number of topics/classes the model really predicted as true [31]:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

where TP and FP represent True Positives and False Positives, respectively.

Recall. Recall computes the proportion of topics/classes that the model properly classified versus all of topics/classes that were real [31]:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

where TP and FN are the true positives and false Negatives respectively.

The F1. We applied the micro $F1$, a micro-averaged $F1$ that is frequently used to assess the prediction performance. The $F1$ score is calculated by averaging Precision and Recall [32]:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

Accuracy. The accuracy is used to measure how many times the classifier predicts correctly. It can be defined as the total number of elements that are predicted correctly (the sum of TP and TN) divided by the total number of predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where TP, FN, FP, and TN are is true positives, false negatives, false positives, and true negatives, respectively.

Table 4. Classification results

Aspect classifier		
	BERT	XLM-RoBERTa
Accuracy	0.72	0.81
Precision	0.67	0.82
Recall	0.67	0.79
$F1$	0.66	0.80
Sentiment classifier		
	BERT	XLM-RoBERTa
Accuracy	0.90	0.94
Precision	0.88	0.92
Recall	0.88	0.93
$F1$	0.88	0.92
Suggestion classifier		
	BERT	XLM-RoBERTa
Accuracy	0.80	0.83
Precision	0.79	0.83
Recall	0.73	0.80
$F1$	0.76	0.81

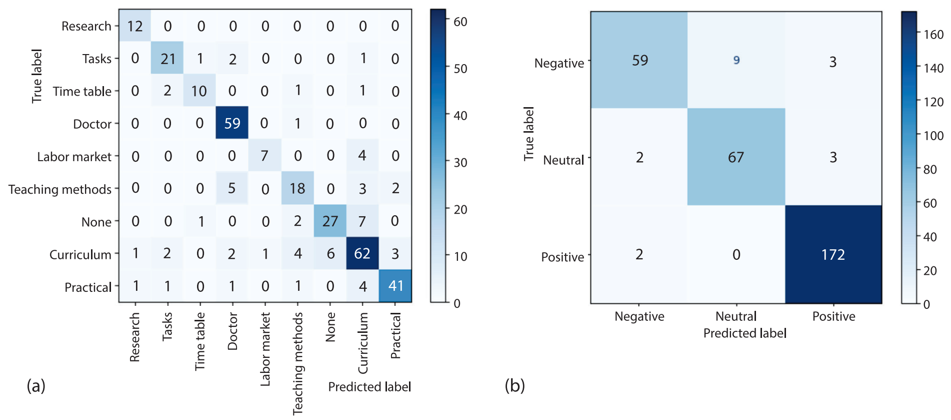
Results and discussion

Aspect classifier. Table 4 presents the findings from the two experiments which show that the XLM-RoBERTa (multilingual) model has demonstrated encouraging results for aspect classification with an $F1$ score of 0.80, precision of 0.82, recall of 0.79 and accuracy of 0.81 compared to BERT multilingual base model (cased). Figure 4(a) shows the confusion matrix of XLM-RoBERTa model, which is a common tool used to evaluate the performance of a classifier [33]. Overall, he model seems to perform well on *Doctor* and *Curriculum*, with relatively high numbers on the diagonal. There are some misclassifications in other classes, such as *Time Table* and *Labor Market*, indicating areas where the model might need improvement.

Sentiment classifier. The XLM-RoBERTa outperforms BERT across all metrics, accuracy, precision, recall, and $F1$ as seen in tab. 4. Figure 4(b) shows XLM-RoBERTa results per sentiment category. It suggests that XLM-RoBERTa model performs well across all sentiment categories, particularly in identifying positive sentiment, though there might be some focus needed on improving the detection of Negative sentiment.

Suggestion classifier. Regarding the suggestion classification, tab. 4 indicates the differences in metrics between the two models and shows that XLM-RoBERTa is more capable of accurately predicting the correct labels, with fewer false positives and false negatives compared to BERT. XLM-RoBERTa performance per label is shown in fig. 4(c). The model performs best in the "*needs to* be added category, with high precision, recall, and $F1$, indicating that the model can identify and predict when something needs to be added.

Table 5 represents some examples of the final output of an ABSA pipeline applied to reviews in the educational domain. Each student's review is breaking down into specific aspects (topics), identifying the sentiment expressed about each aspect, and suggesting potential suggestion based on the sentiment and content of the review. Figure 5 provides the distribution of sentiments across different topics within the course evaluation survey. Overall, the highest positive sentiment is associated with the *Doctor* and *Curriculum* topics, indicating that these aspects are particularly well-received. While fig. 6 shows the distribution of different types of suggestions across various topics. Needs enhancements is commonly suggested across multiple topics indicating areas where improvements are necessary while needs to be removed is less common.



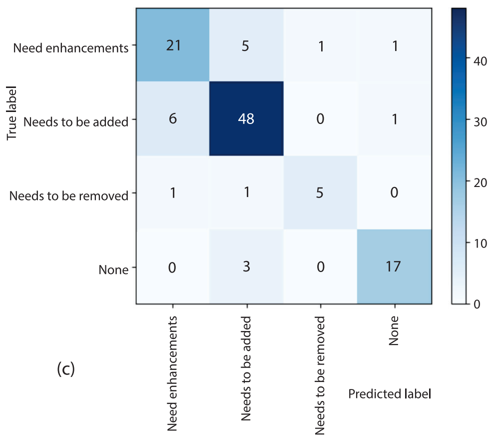


Figure 4. The XLM-RoBERTa confusion matrix;
(a) aspect classifier,
(b) sentiment classifier, and
(c) suggestion classifier

Table 5. Some examples of the final output using XLM-RoBERTa

Review	Topic	Sentiment	Suggestion
Dr is a professional in everything (explaining, organizing, supporting, helping...*etc*)	Doctor	Positive	None
Giving small projects instead of a large project	Tasks	Neutral	Needs enhancements
Integrate the curriculum content with what we take as computer science students in other subjects	Tasks	Neutral	Needs to be added
I particularly enjoyed the teaching used with us. The real-world applications allowed me to apply theoretical concepts to concrete problems	Teaching methods	Positive	None

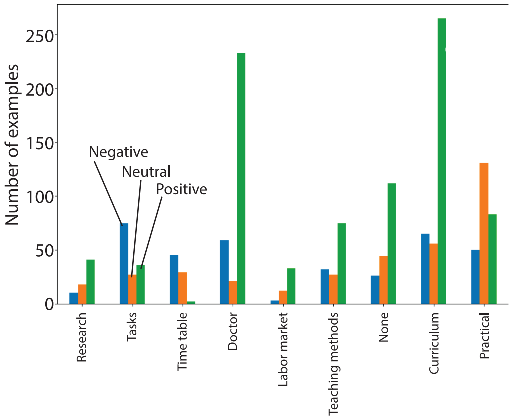


Figure 5. Sentiment count by topic using XLM-RoBERTa

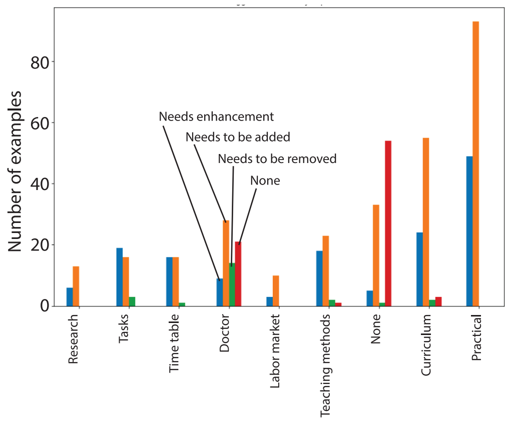


Figure 6. Suggestion count by topic using XLM-RoBERTa

Conclusions

This paper effectively illustrates how ABSA may be applied to open-ended responses in student course evaluation surveys within the context of Arabic education. Utilizing sophisticated transformer-based models like XLM-RoBERTa and BERT, the models achieved significant results in both aspect extraction and sentiment classification. The multilingual XLM-RoBERTa model has shown better performance compared with BERT model, especially in aspect classification.

The results demonstrate how ABSA can give educational institutions more in-depth understanding of students' experiences than can be obtained from closed-ended survey questions. This approach facilitates data-driven decision-making to enhance overall student experiences such as curriculum, instructors, and teaching methods. Furthermore, this study establishes the framework for further research in this field and contributes to the literature on ABSA in Arabic, particularly in educational contexts.

In conclusion, using ABSA in educational surveys can greatly increase the awareness of educational institutions to meet the requirements of their students. It also helps to understand not only how students feel about different aspects of their educational experience but also what actions could be taken to enhance or maintain quality. This kind of analysis is crucial for continuous improvement in educational settings.

References

[1] Bell, A., Chris Brooks, C., What Makes Students Satisfied, A Discussion and Analysis of The Uk's National Student Survey, *Journal of Further and Higher Education*, *42* (2018), 8, pp. 1118-1142
[2] Mizrachi, D., *et al.*, Academic Reading Format Preferences and Behaviors among University Students World-Wide: A Comparative Survey Analysis, *PLOS ONE*, *13* (2018), 05, pp. 1-32
[3] Marshall, P., Contribution of Open-Ended Questions in Student Evaluation of Teaching, *Higher Edu-Cation Research and Development*, *41* (2022), 6, pp. 1992-2005
[4] Baburajan, V., *et al*., Open-Ended Versus Closed- Ended Responses: A Comparison Study Using Topic Modeling and Factor Analysis, *IEEE Tran. on Inte. Trans. Systems*, *22* (2021), 4, pp. 2123-2132
[5] Brauwers, G., Flavius Frasincar, F., A Survey on Aspect-Based Sentiment Classification, *ACM Computing Survey*, 55 (2021), 4
[6] Sun, Q., *et al*., Exploring Ewom in Online Customer Reviews: Sen-Timent Analysis at a Fine-Grained Level, *Engineering Applications of Artificial Intelligence*, *81* (2019), May, pp. 68-78
[7] Shaik, T., *et al*., Sentiment Analysis and Opinion Mining on Educational Data: A Survey, *Natural Language Processing Journal*, *2* (2023), 100003

[8] Zhang, W., *et al*., A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges, *IEEE Transactions on Knowledge and Data Engineering*, *35* (2023), 11, pp. 11019-11038
[9] Mowlaei, M. E., *et al*., Aspect-Based Sentiment Analysis Using Adaptive Aspect-Based Lexicons, *Expert Systems with Applications*, *148* (2022), 113234
[10] Casola, S., *et al*., Pre-Trained Transformers: An Empirical Comparison, *Machine Learning with Applications*, *9* (2022), 100334
[11] Liu, P., *et al*., Pretrain, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, *ACM Computing Surveys*, 55 (2023), 9, pp. 1-35
[12] Sindhu, I., *et al*., Aspect-Based Opinion Mining on Student's Feedback for Faculty Teaching Performance Evaluation, *IEEE Access*, *7* (2019), July, pp. 108729-108741
[13] Kastrati, Z., *et a*l., Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of Moocs, *IEEE Access*, *8* (2020), June, pp. 106799-106810
[14] Sultani M., Daneshpour, N., Extracting Urgent Questions from Mooc Discussions: A Bertbased Multi-Output Classification Approach, *Arabian Journal for Science and Engineering*, On-line first, https://doi.org/10.1007/s13369-024-09090-7, 2024
[15] Hussain, S., *et al*., Aspect2labels: A Novelistic Decision Support System for Higher Educational Institutions by Using Multi-Layer Topic Modelling Approach, *Expert Systems with Applications*, *209* (2022), 118119
[16] Ren, R., *et al*., Automatic Scoring of Student Feedback for Teaching Evaluation Based on Aspect-Level Sentiment Analysis, *Education and Information Technologies*, *28* (2022), July, pp. 797-814
[17] Rogers, A., *et al*., A primer in BERTology: What We Know about How BERT Works, *Transactions of the Association for Computational Linguistics*, *8* (2020), Dec., pp. 842-866
[18] Zhang, H., *et al.*, A Bert Fine-Tuning Model for Targeted Sentiment Analysis of Chinese Online Course Reviews, *International Journal on Artificial Intelligence Tools*, *29* (2020), 07-08, 2040018
[19] Alshaikh, K. A., *et al*., *et al.*, Bert-Based Model for Aspect-Based Sentiment Analysis for Analyzing Arabic open-Ended Survey Responses: A Case Study, *IEEE Access*, *12* (2024), Dec., pp. 2288-2302
[20] Mustafa, M., *et al*., Arabic Aspect Sentiment Polarity Classification Using Bert, *Journal of Big Data*, *9* (2022)*,* 12
[21] Chennafi, M. E., *et al*., Arabic Aspect-Based Sentiment Classification Using Seq2Seq Dialect Normalization and Transformers, *Knowledge*, *2* (2022), 3, pp. 388-401
[22] AlJarrah, I., *et al*., Aspect-Based Sentiment Analysis for Arabic Food Delivery Reviews, *ACM Transactions on Asian and Low-Resource Language Information Processing*, *22* (2023), July, pp. 2375-4699
[23] Hasna Chouikhi, H., Mohammed Alsuhaibani, and Fethi Jarray. Bertbased Joint Model for Aspect Term Extraction and Aspect Polarity Detection in Arabic Text, *Electronic*s, *12* (2023), 3, pp. 2079-9292
[24] Almasri, M., *et al*., A Semi Supervised Approach to Arabic Aspect Category Detection Using Bert and Teacher-Student Model, *PeerJ Computer Science*, *9* (2023), e1425
[25] Bensoltane R., Zaki, T., Towards Arabic Aspect-Based Sentiment Analysis: A Transfer Learning-Based Approach, *Social Network Analysis and Mining*, *12* (2021), Nov., pp. 1-16
[26] El-Alami, F. Z., *et al*., Contextual Semantic Embeddings Based on Fine-Tuned Arabert Model for Arabic Text Multi-Class Categorization, *Journal of King Saud University-Computer and Information Sciences*, *34* (2022), 10, pp. 8422-8428
[27] Li, B., *et al*., Robust Dimensionality Reduction Via Feature Space to Feature Space Distance Metric Learning, *Neural Networks, 11* (2019), 2, pp. 1-14
[28] Campello, R. J. G. B., *et al*., *Data Mining and Knowledge Discovery*, *10* (2020), 2, e1343
[29] George, L., Sumathy, P., An Integrated Clustering and Bert Framework for Improved Topic Modeling *International Journal of Information Technology*, *15* (2023), 4, pp. 2187-2195
[30] Abuzayed, A., Al-Khalifa, H., Bert for Arabic Topic Modeling: An Experimental Study on Bertopic Technique, *Procedia Computer Science*, *189* (2021), 1, pp. 191-194
[31] Miao, J., Zhu, W., Precision-Recall Curve (PRC) Classification Trees, *Evolutionary Intelligence*, *15* (2022), 09
[32] Christen, P., *et al.*, A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives, *ACM Computing Surveys*, *56* (2023), 3, pp. 1-24
[33] Xu, J., *et al*., Three-Way Confusion Matrix for Classification: Ameasure Driven View, *Information Science*s, *507* (2020), Jan., pp. 772-794

Paper submitted: August 3, 2024
Paper revised: October 20, 2024
Paper accepted: November 21, 2024
2024 Published by the Vinča Institute of Nuclear Sciences, Belgrade, Serbia.
This is an open access article distributed under the CC BY-NC-ND 4.0 terms and conditions.

Copyright of Thermal Science is the property of Society of Thermal Engineers of Serbia and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.