Comparative Analysis of Topic Modelling Approaches on Student Feedback

Faiz Hayat[1][a], Safwan Shatnawi[2][b] and Ella Haig[1][c]
[1] *School of Computing, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth PO1 3HE, U.K.*
[2] *Education Practice, International Research and Exchange Board, Amman, Jordan*

Keywords: Topic Modelling, BERT, LDA, LSA, NMF, Education.

Abstract: Topic modelling, a type of clustering for textual data, is a popular method to extract themes from text. Methods such as Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA) and Non-negative Matrix Factorization (NMF) have been successfully used across a wide range of applications. Large Language Models, such as BERT, have led to significant improvements in machine learning tasks for textual data in general, as well as topic modelling, in particular. In this paper, we compare the performance of a BERT-based topic modelling approach with LDA, LSA and NMF on textual feedback from students about their mental health and remote learning experience during the COVID-19 pandemic. While all methods lead to coherent and distinct topics, the BERT-based approach and NMF are able to identify more fine-grained topics. Moreover, while NMF resulted in more detailed topics about the students' mental health-related experiences, the BERT-based approach produced more detailed topics about the students' experiences with remote learning.

1 INTRODUCTION

Machine learning tasks are typically divided into supervised and unsupervised learning (Berry et al., 2019). For textual data, one of the most used unsupervised methods is topic modelling, which is a type of clustering that extracts topics or themes from text (Zhao et al., 2021).

Three of the most popular methods for topic modelling are Latent Dirichlet Allocation (LDA) (Blei et al., 2003), Latent Semantic Analysis (LSA) (Deerwester et al., 1990) and Non-negative Matrix Factorization (NMF) (Lee and Seung, 1999). Since the arrival of Large Language Models (LLMs) in 2017 (Vaswani et al., 2017), pre-trained deep learning models like BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al., 2019) have shown impressive results for unsupervised learning across many applications (e.g., (Abuzayed and Al-Khalifa, 2021; Egger and Yu, 2022; Sharifian-Attar et al., 2022)). Compared with other topic modelling approaches, BERT-based models have the following two key advantages: (1) because they were trained on large amounts of data, they have the capacity to encode complex semantic relationships, and (2) the ability to capture both left and right contexts, which accounts for the term "bidirectional" (Devlin et al., 2019).

In this paper, we compare a BERT-based topic modeling approach with LDA, LSA, and NMF to identify relevant topics from student feedback on their COVID-19 pandemic experience, focusing on mental health and remote learning.

The main contribution of the paper is a comparative analysis of topic modeling using LLMs like BERT against traditional methods. Few studies have explored this comparison, leaving the superiority of newer approaches uncertain. We investigate whether BERT provides an advantage over traditional methods in analyzing student feedback on pandemic experiences. Our study compares topics identified using BERT-based modeling with NMF, LDA, and LSA.

The rest of the paper is structured as follows: Section 2 reviews background and related work, Section 3 details the experimental setup, Section 4 presents the results, Section 5 compares methods and discusses findings and Section 6 concludes with future research directions.

[a] https://orcid.org/0000-0003-0249-4617
[b] https://orcid.org/0000-0002-5063-1295
[c] https://orcid.org/0000-0002-5617-1779

Hayat, F., Shatnawi, S. and Haig, E.
Comparative Analysis of Topic Modelling Approaches on Student Feedback.
DOI: 10.5220/0012890400003838

In *Proceedings of the 16th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2024) - Volume 1: KDIR*, pages 226-233
ISBN: 978-989-758-716-0; ISSN: 2184-3228

2 RELATED WORK

This section provides an overview of LDA, LSA, and NMF, reviews research on student feedback, and discusses evaluation approaches for topic modeling.

LDA is a probabilistic generative model widely used for topic modeling in natural language processing (NLP) (Blei et al., 2003). It assumes that each document in a corpus is a mixture of topics, and each topic is a distribution over words. LDA aims to uncover latent topics from a collection of documents by iteratively assigning words to topics and adjusting assignments to maximize the likelihood of the data.

LSA (Deerwester et al., 1990) is a technique used for dimensionality reduction and semantic analysis of textual data. It employs Singular Value Decomposition (SVD) to identify latent semantic structure in a corpus by capturing relationships between terms and documents, representing them in a lower-dimensional space for easier detection of semantic similarities.

NMF (Lee and Seung, 1999) is a dimensionality reduction technique widely used in natural language processing (NLP) and other fields. It decomposes a non-negative matrix into two lower-dimensional matrices, representing topics and document-topic distributions. NMF is applied to tasks such as topic modeling, document clustering, and feature extraction.

BERT (Grootendorst, 2022) is a pre-trained deep learning model developed by Google for natural language processing tasks. It excels in capturing contextual information bidirectionally, enabling it to understand the meaning of words in context more effectively than previous models. BERT has revolutionized NLP tasks by leveraging large-scale pre-training on vast text data and fine-tuning for specific downstream tasks, e.g., (Ding et al., 2023; Malladi et al., 2023).

Topic modelling has been used to analyse student feedback in many studies, e.g., (Buenano-Fernandez et al., 2020; Hujala et al., 2020; Sun and Yan, 2023). There have also been several studies investigating student experiences during the COVID-19 pandemic (e.g., (Oliveira et al., 2021; Stevanović et al., 2021; Waheeb et al., 2022). Many studies also employ BERT-based models with educational-related data (e.g., (Bai and Stede, 2023; Cochran et al., 2023; Sung et al., 2019)). However, to our knowledge, only one study has used BERT-based topic modeling on student feedback (Masala et al., 2021), and none have focused on students' COVID-19 experiences through open-text responses.

We are aware that other studies (e.g., (Müller et al., 2023; Wang et al., 2020; Xu et al., 2022)) have used BERT-based topic modeling to examine COVID-19 experiences in the general population, but they focus on social media data, not student responses from open-ended questionnaires. Therefore, these studies are not directly relevant to our research.

The one study we found using a BERT-based topic modeling technique (Masala et al., 2021) concentrated on examining student textual feedback at the course level. The researchers developed a tool that analyzed large volumes of student feedback, producing clusters of similar contexts and recurring keywords for each course. The processing pipeline involved extracting general evaluations, restoring diacritics using RoBERT (a Romanian BERT model), and performing keyword extraction with KeyBERT (fine-tuned for the Romanian language). To capture the context around these keywords, they utilized two methods: extracting sentences containing the keywords and using dependency tree traversal to gather related context. The extracted contexts were then grouped using K-Means clustering applied to BERT-generated embeddings.

In contrast, our study applies multiple topic modeling techniques to analyze survey responses related to mental health and remote learning during the COVID-19 pandemic. We explore several algorithms, including BERT-based embeddings, Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA), and Non-negative Matrix Factorization (NMF). Additionally, our study incorporates dimensionality reduction using UMAP and clustering with HDBSCAN to discover underlying topics in the survey data.

We now turn our attention to the evaluation of topic modeling techniques. In supervised learning, techniques are evaluated by comparing the predictions against a known ground truth, but in unsupervised learning, such ground truth is often absent, making evaluation challenging without human judgment. Although various metrics are used to evaluate topic modeling and clustering methods, performance can vary widely across techniques and data types (Doogan and Buntine, 2021; Harrando et al., 2021), and the validity of fully automated evaluations without human judgment has been questioned (Hoyle et al., 2021).

Some clustering/topic modelling techniques require as input the number of clusters/topics, while for others, the 'optimal' number emerges from the data. For the former, metrics like coherence scores (Abdelrazek et al., 2023; O'Callaghan et al., 2015), can help determine the optimal number of topics, but these also need human judgment (Doogan and Buntine, 2021). In our research, we combined coherence scores with human evaluation.

While the usefulness of BERT-based approaches for topic modelling has been shown for different types of education-related data, there has only been one study using a BERT-based approach on student feedback from open-ended questions and this study did not include a comparison with other topic modelling approaches. Our study contributes to a better understanding of the usefulness of BERT by providing the first comparative study for this type of data.

3 EXPERIMENTAL SETUP

In this section, we describe the data collection and preprocessing, as well as the process for topic modelling for each of the four investigated approaches.

3.1 Data Collection

Data collection for this research involved conducting a survey among students at a UK university in 2022. The aim was to assess the influence of the COVID-19 pandemic on students. The questionnaire included four open-ended prompts designed to clarify the particular difficulties students encountered regarding their mental well-being and remote learning during the pandemic: 'What challenges or issues regarding mental health did you face during the pandemic? What aspects, if any, did you struggle with?'; 'Please share any other comments/ opinions/ solutions about your mental health during the pandemic.'; 'What challenges or issues regarding remote learning did you face during the pandemic? What aspects, if any, did you struggle with?' and 'Please share any other comments/ opinions/ solutions about remote learning during the pandemic.'

Ethical approval was obtained from the university's Ethics Committee before distribution. The survey was distributed using email lists specific to each faculty, reaching out to a diverse group of students from different academic disciplines such as social sciences, humanities, business and law, and technology. The involvement in the survey was voluntary and respondents remained anonymous.

Responses from 340 participants included 696 submissions from the open-ended questions: 375 on mental health and 321 on remote learning. The sample size for topic modeling consisted of all 696 textual responses. We made this decision due to the prevalence of short answers and many students responding selectively to some questions and not others.

3.2 Data Preprocessing

Data preprocessing was conducted to prepare the textual data for analysis. Specifically, this process included the elimination of stop words, such as "the," "is," and "and", which are common words that provide little value in understanding the underlying themes of the text. Special characters and numbers not contributing to semantic analysis were also filtered out to refine the dataset and improve the quality of information fed into the topic modeling algorithm.

3.3 Topic Modeling Algorithms

Four topic modeling algorithms were utilized: a BERT approach described below, Latent Dirichlet Allocation (LDA), Latent Semantic Analysis (LSA) and Non-negative Matrix Factorization (NMF). The implementation was carried out using Google Colab, a cloud-based environment integrated with Python.

The BERT-Based Topic Modelling Approach. The following steps were applied: 1) Obtaining document embeddings by utilizing the 'paraphrase-MiniLM-L6-v2' pre-trained model; 2) UMAP (Uniform Manifold Approximation and Projection) (McInnes et al., 2018) was used to reduce the dimensionality of the embeddings, improving visualization and clustering; 3) Performing clustering using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al., 2017) algorithm to generate the topics; 4) Using visualizations to analyze the 10 most frequent words per topic, response distribution across topics, and the dendrogram from the clustering algorithm; 5) Conducting qualitative analysis to validate topics by examining responses assigned to each topic.

Latent Dirichlet Allocation (LDA). The following steps were applied: 1) Using the Gensim library, a dictionary and a document-term matrix were created to represent the term frequency; 2) Applied LDA to the document-term matrix to infer the underlying topics and their word distributions; 3) Analyzing the resulting topics by examining the most probable words associated with each topic; 4) Conducting qualitative analysis to validate the topics by reviewing documents assigned to each topic.

LSA (Latent Semantic Analysis). The following steps were applied to derive the topics using LSA: 1) Creating a term-document matrix representing the frequency of terms in documents. 2) Applying SVD to the term-document matrix to decompose it into three matrices: a term-concept matrix, a diagonal matrix of singular values, and a concept-document matrix. 3) Analyzing the resulting concept vectors to identify latent semantic topics. 4) Conducting qualitative analysis to validate the topics by reviewing documents associated with each concept.

NMF (Non-Negative Matrix Factorization). The following steps were applied: 1) Vectorizing the preprocessed text data into a term-document matrix, where each row represents a document and each column represents a term. 2) Applying NMF to factorize the term-document matrix into two matrices representing topics and document-topic distributions. 3) Analyzing the resulting topics by examining the most prominent terms associated with each topic. 4) Conducting qualitative analysis to validate the topics by reviewing documents assigned to each topic.

4 RESULTS

As the BERT-based approach uses the HDBSCAN algorithm, the optimal number of topics emerges from the data; in our case, this was 13. LDA, LSA and NMF require a number of topics as an input. For these methods, to identify the optimal number of topics, as mentioned in Section 2, we chose the coherence score (Abdelrazek et al., 2023), which aggregates the coherence of each topic, measured as the semantic similarity between top words in the topic, in combination with human judgment. The highest coherence scores were obtained for 13 topics with LDA, 12 topics with LSA, and 16 topics with NMF, and our qualitative evaluation showed that for each method the topics were relevant and distinct from each other.

We conducted a deeper qualitative assessment of topics from all four algorithms and found the BERT-based approach and NMF yielded the most interesting results. Due to space constraints, we present detailed results for these methods and summarize LDA and LSA results for comparison in the next section.

The topics that resulted from the BERT-based approach are presented in Table 1. We grouped the topics into themes, analyzed in the following paragraphs.

As anticipated, we see that the subjects are arranged in relation to the two elements—mental health and remote learning—that were highlighted in the open-ended questions. Out of the thirteen topics, three (0 and 4-5) are related to mental health, two are related to both (1 and 6) and eight topics (2-3 and 7-12) are related to distant learning.

The application of the BERT-based modeling approach to mental health allows differentiation between several aspects, including anxiety (Topic 0), social isolation and loneliness (Topic 4), and the generic impact of the epidemic on mental health (Topic 5).

It is interesting to note that a more comprehensive picture of remote learning emerges, covering a wide range of topics, from the more general ones like the university experience in general (Topic 3) and the impact of the pandemic on the university experience (Topic 8), to the more specialised ones like concentration problems (Topic 2), internet connectivity (Topic 7), virtual communication (Topic 9), lecture formats (Topic 10), the experience of remote learning across various modules and courses (Topic 11), and the value of in-person communication (Topic 12).

Aspects of both remote learning and mental health are included in Topics 1 and 6. In Topic 1, motivation is discussed as a practical requirement for participating in remote learning, as well as a crucial component of mental health. The only positive topic is Topic 6, which describes the methods respondents use to preserve their mental health and academic motivation.

Table 1 displayed the number of textual instances per topic in the second column, with a relatively large variation. Topic 12 (face-to-face communication) has the fewest instances (14), while Topic 5 (the pandemic's effects on mental health) has the highest (88).

There are parallels between Topics 5 and 8, which discuss how the pandemic has affected mental health (Topic 5) and remote learning (Topic 8), respectively. Topics 10 and 11 share commonalities as well, as they both deal with challenges related to remote learning. The variations between the two topics highlight experiences related to lectures and teaching sessions in Topic 10 and broader experiences related to remote learning at the module or course level in Topic 11.

As mentioned in Section 3.1, responses to mental health (375) outnumbered those to distant learning (321). The fact that there are three topics about mental health and eight about remote learning suggests that while there are more different experiences with remote learning, there is a greater homogeneity of experiences with mental health. This further demonstrates the capacity of the BERT-based approach to discern between elements with subtle variations.

Table 2 presents the topics resulting from applying NMF. Similar to the BERT-based approach, the topics cover mental health, remote learning, or both aspects.

In terms of mental health, specific issues such as anxiety, eating disorders and depression are covered in Topic 1, dealing with uncertainty in Topic 3, and the generic impact of the pandemic on mental health in Topic 8. Topic 10 is also more generic, covering emotional well-being aspects, while Topic 11 is more specifically about social isolation challenges.

The topics covering remote learning aspects vary from more generic, about distance learning and the use of online tools (Topics 2, 6, and 9), to more specific issues such as motivation to study (Topic 12) and difficulties in grasping learning content (Topic 13).

Several topics cover both mental health and remote learning aspects: time management (Topic 0), motivational issues (Topics 4 and 15), lack of social interaction in remote learning (Topic 5), the impact of the pandemic on physical health (Topic 7), and the

Table 1: Topics extracted with the BERT-based approach; (Docs refers to the number of documents/responses for each topic).

No.	Docs	Topic Name	Topic Description	Keywords
0	34	Anxiety and Depression	Increased anxiety and depression disorders, leading to heightened awareness and impacts on mental health.	anxiety, depression, disorders, increased, eating, panic, depressed, still, health, aware
1	34	Motivation	Struggle to maintain motivation, resulting in challenges in staying focused and productive.	motivation, motivated, stay, keep, staying, lack, work, hard, struggle
2	26	Easily Distracted	Experience difficulty concentrating due to various distractions, affecting productivity and focus.	distracted, easily, concentrate, focused, couldnt, distractions, attention, skip, work, focus
3	19	University Experience	Mixed experiences during university, including success, failure, and uncertainty	university, year, felt, experience, well, think, failed, cheated, second, uni
4	74	Loneliness and Friendship	Loneliness and lack of social contact affecting mental well-being and interaction.	friends, loneliness, lonely, social, see, depression, lack, contact, isolation, able
5	88	Pandemic and Mental Health	Heightened awareness of mental health issues during the pandemic, affecting individuals and communities globally.	pandemic, health, mental, people, social, covid, anxiety, family, made, measures
6	19	Daily Routine	Recognizing the value of a consistent, positive daily routine for better mental health.	daily, good, routine, mental, health, work, home, day, weekly, sleep
7	16	Internet Connection Issues	Frustration and challenges from unreliable internet, affecting academic and personal tasks.	internet, connection, bad, unreliable, issues, lesson, poor, exams, could, found
8	27	Remote Learning	Adapting to remote learning challenges, including online lectures and assignments.	pandemic, remote, learning, working, lectures, really, time, away, home, im
9	19	Zoom Calls	Adjusting to the challenges and discomfort associated with online video calls, especially in educational and professional settings.	zoom, camera, calls, people, would, anyone, lessons, interacting, comfortable, answer
10	50	Online Lectures	Facing challenges with online lectures, including slower learning and engagement issues.	lectures, lecturers, questions, online, without, lecture, students, felt, slower
11	52	Remote Learning Experience	Reflecting on remote learning, its benefits, and drawbacks compared to traditional methods.	learning, remote, time, modules, learn, teachers, students, like, lectures, course
12	14	Face-to-Face Learning	Emphasizing the importance of face-to-face interaction in learning environments for effective communication and understanding.	face, union, learning, lower, guidance, lecturers, communication, important, facetoface, seeing

need for support during studies (Topic 14).

The distribution of responses per topic, unlike BERT-based approaches, NMF has a more balanced range, with the smallest topic having 25 responses (Topic 3) and the largest 60 (Topic 14). Eleven of the sixteen topics have between 40 and 50 responses.

5 COMPARISON AND DISCUSSION

To compare the four algorithms, we selected four themes that cover all the topics produced across all four solutions: remote learning and challenges, mental health and challenges, social issues and loneliness, and motivation and physical health. The topic distribution by theme is shown in Table 3, and Fig. 1 illustrates the theme proportions for each algorithm.

The BERT-based approach allocates the highest percentage (61.54%) of its thematic content to Remote Learning and Challenges, indicating its strong emphasis on analyzing issues related to remote education. Conversely, it allocates smaller proportions

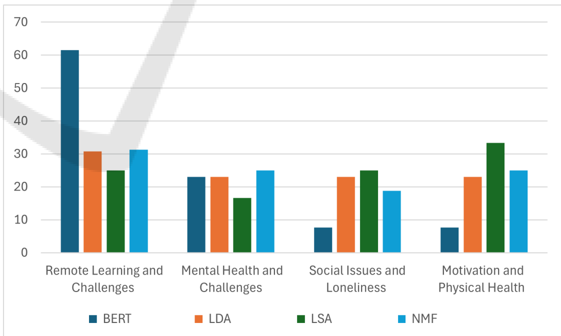

Figure 1: Comparitive Analysis based on Themes.

to Mental Health and Challenges (23.08%), Social Issues and Loneliness (7.69%), and Motivation and Physical Health Issues (7.69%), suggesting a relatively narrower focus on these domains.

LDA (Latent Dirichlet Allocation) has the highest percentage to Remote Learning and Challenges (30.77%), with smaller proportions for Mental Health and Challenges, Social Issues and Loneliness, and Motivation and Physical Health, each at (23.08%).

Table 2: Topics extracted using NMF.

No.	Docs	Topic Name	Topic Description	Keywords
0	59	Time Management Struggles	Focuses on time management challenges worsened by pandemic-related work-life disruptions	struggled, focus, working, time, work, helped, day, home, pandemic, lot
1	30	Mental Health Challenges	Addresses severe mental health issues, including heightened levels of depression and anxiety due to societal pressures.	severe, reached, leaving, eating, society, disorder, house, increased, depression, anxiety
2	50	Remote Learning Preferences	Highlights difficulties adapting to remote learning and a preference for traditional face-to-face interactions.	difficulty, pace, better, lecture, preferred, tutor, prefer, remote, face, learning
3	25	Coping with Uncertainty	Discusses struggles with coping mechanisms during times of uncertainty, leading to feelings of loneliness and boredom.	uncertainty, email, change, coping, boring, struggling, extremely, covid, help, loneliness
4	49	Motivation Struggles	Focuses on maintaining motivation for completing coursework, with challenges in maintaining consistent effort.	module, getting, far, complete, week, whilst, went, struggled, motivation, work
5	40	Lack of Social Interaction	Explores the absence of social interactions in learning environments, leading to feelings of disconnection.	teacher, medium, unable, aspect, make, seeing, talking, people, interaction, social
6	40	Challenges with Distance Learning	Addresses difficulties in maintaining engagement and interaction in distance learning settings.	contact, distance, interaction, long, learning, student, issue, lecture, online, lack
7	45	Impact on Physical Health	Examines how disrupted routines and less exercise affected health during the pandemic.	low, exercise, issue, daily, pandemic, struggle, routine, good, mental, health
8	43	Impact on Mental Health	Examines worsening mental health from isolation, academic stress, and future uncertainty.	worse, life, depression, parent, job, caused, worried, stress, isolation, feel
9	45	Online Learning Experience	Evaluate online learning tools like Zoom, highlighting effectiveness and engagement issues.	use, useful, attention, session, zoom, know, people, lecture, online, class
10	50	Emotional Well-being	Addresses emotional challenges during university, such as stress, depression, and loneliness.	quite, stressed, teaching, university, feel, year, depressed, lonely, like, felt
11	40	Social Isolation Challenges	Explores challenges in maintaining social connections with family and friends due to prolonged social isolation.	future, kept, knowing, member, socialise, difficult, person, able, family, friend
12	35	Study Motivation	Discusses maintaining study motivation and focus amid distractions and coursework demands.	skill, studying, lesson, money, course, focused, stay, staying, motivated, hard
13	45	Understanding Course Material	Explores difficulties in grasping course material, especially under the distractions and pressures of lockdowns.	happened, understanding, lockdown, grade, thing, losing, understand, study, assignment, time
14	60	Academic and Financial Challenges	Addresses challenges in academics and financial stability, highlighting the need for institutional and peer support.	needed, course, poor, socialising, people, financial, harder, really, lecturer, support
15	40	Exam Preparation Challenges	Discusses challenges in preparing for exams due to distractions and unreliable internet connections.	exam, concentrate, difficult, bad, especially, learn, distracted, connection, easily, internet

In contrast, LSA (Latent Semantic Analysis) produces a unique thematic distribution in comparison with the other methods. It assigns a substantially higher percentage (33.34%) to Physical Health Issues and Motivation, indicating a strong emphasis on these two areas. There is no difference between Social issues and Loneliness (25%) and Remote Learning and Challenges(25%). It does, however, give Mental Health and Challenges a lower percentage (16.67%).

NMF allocated the highest percentage (31.25%) to Remote Learning and Challenges, equal percentages to Mental Health and Challenges, and Motivation and Physical Health (23.08%), and the lowest percentage (18.75%) to Social Issues and Loneliness.

Overall, NMF and LDA have the most balanced distributions across the four themes and can capture

Table 3: Algorithmic Topics Distribution.

Algorithm	Remote Learning and Challenges	Mental Health and Challenges	Social Issues and Loneliness	Motivation and Physical Health
BERT	Topic 2, 3, 7, 8, 9, 10, 11, 12	Topic 0, 5, 6	Topic 4	Topic 1
LDA	Topic 2, 8, 9, 12	Topic 0, 4, 6	Topic 3, 5, 10	Topic 1, 7, 11
LSA	Topic 3, 5, 7	Topic 0, 8	Topic 1, 2, 9	Topic 4, 6, 10, 11
NMF	Topic 0, 2, 6, 9, 13	Topic 1, 3, 8, 10	Topic 5, 11, 14	Topic 4, 7, 12, 15

at a good level of detail several distinct aspects. The BERT-based approach, on the other hand, has a more unbalanced distribution across the four themes but can capture more fine-grained issues related to remote learning. In particular, three topics identified by the BERT-based approach were not identified as separate topics by any of the other algorithms: internet connection issues (Topic 7), online calls (Topic 9), and face-to-face learning (Topic 12). By volume of responses, these are also among the smallest topics, with 16, 19 and 14 responses, respectively. From this point of view, the BERT-based approach may be better when a more fine-grained picture would be of interest.

For all approaches, we applied data preprocessing, as outlined in Section 3.2. There is very little empirical evidence concerning the use of textual data preprocessing when pre-trained LLMs are used. We applied the BERT-based approach with no preprocessing as well as the preprocessing mentioned in Section 3.2 and found more coherent results when using preprocessing, hence, we reported the results with preprocessing. This aligns with the view that preprocessing should still be considered for LLMs expressed in a recent review of text preprocessing (Chai, 2023).

6 CONCLUSION

This paper presents a comparative study using four topic modeling methods: BERT, LDA, LSA, and NMF, on student feedback in textual format about the mental health and remote learning students' experiences during the COVID-19 pandemic.

This study sought to determine the effectiveness of the BERT topic model compared to traditional approaches like NMF, LDA, and LSA. The results indicated that BERT provided deeper insights into remote learning challenges during the pandemic. While traditional methods produced similar results in mental health, social issues, isolation, and motivation, BERT showed clear advantages in topic understanding.

Our study found that all methods produced coherent topics covering various aspects, but BERT and NMF generated more interesting topics than LDA and LSA. NMF had a balanced response distribution, while BERT exhibited significant variation.

The two primary limitations of our study are the sample size and the post-epidemic data collection, which may have influenced students' recollections. We gathered 696 textual instances from 340 participants. Despite this small sample, all algorithms produced coherent topics.

Among the four algorithms, the BERT-based approach was least affected by the small sample size due to its extensive pre-training, which may explain its ability to capture more nuanced topics. Our research highlights the potential of BERT-based topic modeling for educational data. In the future we will explore alternative BERT models, like DeBERTa (He et al., 2020), known for its effectiveness in textual emotion recognition (Boitel et al., 2023), to capture more emotionally nuanced experiences.

REFERENCES

Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., and Hassan, A. (2023). Topic modeling algorithms and applications: A survey. *Information Systems*, 112:102131.

Abuzayed, A. and Al-Khalifa, H. (2021). BERT for arabic topic modeling: An experimental study on BERTopic technique. *Procedia computer science*, 189:191–194.

Bai, X. and Stede, M. (2023). A survey of current machine learning approaches to student free-text evaluation for intelligent tutoring. *International Journal of Artificial Intelligence in Education*, 33(4):992–1030.

Berry, M. W., Mohamed, A., and Yap, B. W. (2019). *Supervised and unsupervised learning for data science*. Springer.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(4-5):993–1022.

Boitel, E., Mohasseb, A., and Haig, E. (2023). A comparative analysis of GPT-3 and BERT models for text-based emotion recognition: performance, efficiency, and robustness. In *UK Workshop on Computational Intelligence*, pages 567–579. Springer.

Buenano-Fernandez, D., Gonzalez, M., Gil, D., and Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access*, 8:35318–35330.

Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3):509–553.

Cochran, K., Cohn, C., Hastings, P., Tomuro, N., and Hughes, S. (2023). Using BERT to identify causal structure in students' scientific explanations. *Artificial Intelligence in Education*, pages 1–39.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *The Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, page 4171–4186.

Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., Hu, S., Chen, Y., Chan, C.-M., Chen, W., et al. (2023). Parameter-efficient fine-tuning of large-scale pre-trained language models. *Nature Machine Intelligence*, 5(3):220–235.

Doogan, C. and Buntine, W. (2021). Topic Model or Topic Twaddle? Re-evaluating Semantic Interpretability Measures. In *North American Association for Computational Linguistics 2021*, pages 3824–3848.

Egger, R. and Yu, J. (2022). A topic modelling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts. *Frontiers in Sociology*, 7:886498.

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

Harrando, I., Lisena, P., and Troncy, R. (2021). Apples to apples: A systematic evaluation of topic models. In *The International Conference on Recent Advances in Natural Language Processing*, pages 483–493.

He, P., Liu, X., Gao, J., and Chen, W. (2020). DeBERTa: Decoding-enhanced BERT with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Hoyle, A., Goel, P., Hian-Cheong, A., Peskov, D., Boyd-Graber, J., and Resnik, P. (2021). Is Automated Topic Model Evaluation Broken?: The Incoherence of Coherence. *Advances in neural information processing systems*, 34:2018–2033.

Hujala, M., Knutas, A., Hynninen, T., and Arminen, H. (2020). Improving the quality of teaching by utilising written student feedback: A streamlined process. *Computers & Education*, 157:103965.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791.

Malladi, S., Gao, T., Nichani, E., Damian, A., Lee, J. D., Chen, D., and Arora, S. (2023). Fine-tuning language models with just forward passes. In *Advances in Neural Information Processing Systems*, volume 36, pages 53038–53075.

Masala, M., Ruseti, S., Dascalu, M., and Dobre, C. (2021). Extracting and clustering main ideas from student feedback using language models. In *International Conference on Artificial Intelligence in Education*, pages 282–292. Springer.

McInnes, L., Healy, J., Astels, S., et al. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11):205.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Müller, M., Salathé, M., and Kummervold, P. E. (2023). COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter. *Frontiers in artificial intelligence*, 6:1023281.

Oliveira, G., Grenha Teixeira, J., Torres, A., and Morais, C. (2021). An exploratory study on the emergency remote education experience of higher education students and teachers during the COVID-19 pandemic. *British Journal of Educational Technology*, 52(4):1357–1376.

O'Callaghan, D., Greene, D., Carthy, J., and Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, 42(13):5645–5657.

Sharifian-Attar, V., De, S., Jabbari, S., Li, J., Moss, H., and Johnson, J. (2022). Analysing longitudinal social science questionnaires: Topic modelling with BERT-based embeddings. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 5558–5567.

Stevanović, A., Božić, R., and Radović, S. (2021). Higher education students' experiences and opinion about distance learning during the Covid-19 pandemic. *Journal of Computer Assisted Learning*, 37(6):1682–1693.

Sun, J. and Yan, L. (2023). Using topic modeling to understand comments in student evaluations of teaching. *Discover Education*, 2:1–12.

Sung, C., Dhamecha, T. I., and Mukhi, N. (2019). Improving short answer grading using transformer-based pretraining. In *International Conference on Artificial Intelligence in Education*, pages 469–481.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Waheeb, S. A., Khan, N. A., and Shang, X. (2022). Topic modelling and sentiment analysis of online education in the COVID-19 era using social networks-based datasets. *Electronics*, 11(5):715.

Wang, T., Lu, K., Chow, K. P., and Zhu, Q. (2020). COVID-19 sensing: negative sentiment analysis on social media in china via BERT model. *IEEE Access*, 8:138162–138169.

Xu, W. W., Tshimula, J. M., Dubé, É., Graham, J. E., Greyson, D., MacDonald, N. E., and Meyer, S. B. (2022). Unmasking the Twitter discourses on masks during the COVID-19 pandemic: User cluster–based BERT topic modeling approach. *JMIR Infodemiology*, 2(2):e41198.

Zhao, H., Phung, D., Huynh, V., Jin, Y., Du, L., and Buntine, W. (2021). Topic modelling meets deep neural networks: a survey. In *The AAAI International Joint Conference on Artificial Intelligence 2021*, pages 4713–4720.