Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 19 No. 16 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijim.v19i16.56253

PAPER

Detection of Cognitive Distortions in Students' Thoughts Using Topic Modeling and Fuzzy Clustering

Marwah Alian[1,2] (✉), Maisa Al-Khazaleh[2], Hadeel AlShboul[1]

[1]Computer Science Department, Faculty of Information Technology, The World Islamic Sciences and Education University, Amman, Jordan

[2]Basic Sciences Department, Faculty of Science, The Hashemite University, Zarqa, Jordan

marwah.alian@wise.edu.jo

ABSTRACT

Cognitive distortion (CD) refers to an irrational thinking pattern that causes individuals to misinterpret information and convince themselves of incorrect information. This study investigates patterns of CDs in students' thoughts after academic exams. Machine learning models are utilized to detect and categorize these distortions. The methodology of this study utilizes topic modeling with two approaches: latent Dirichlet allocation (LDA) with Tf–Idf and non-negative matrix factorization (NMF). The NMF approach is applied with two different pre-trained embeddings (AraBERT and AraGPT). Fuzzy clustering is combined with these topic modeling approaches, and the results are compared. Experiments are conducted using two datasets: a collection of students' thoughts and a generated dataset that is based on cognitive behavioral therapy (CBT) principles. When analyzing the students' thoughts dataset, NMF with AraBERT demonstrated superior performance by producing the most meaningful topics with a coherence score of 0.78. However, in the generated dataset, NMF with AraGPT achieved a better balance between coherence and separation, along with clearer topic boundaries. Although NMF with AraBERT achieves the highest coherence score (0.86), it shows significant topic overlap inferred from the inter-clustering score (0.81). Fuzzy clustering, topic modeling, and NMF-AraGPT together provide the highest overall performance when applied to the students' dataset. This combination provides distinct and well-separated topics inferred from the inter-clustering score (0.53). NMF topic modeling with AraGPT is the most effective model when integrated with fuzzy clustering based on the comprehensive analysis.

KEYWORDS

cognitive distortion (CD), fuzzy clustering, topic modeling, AraGPT, AraBERT

1 INTRODUCTION

Cognitive behavioral therapy (CBT) is a structured therapeutic approach of psychotherapy that focuses on the relationship between thoughts, feelings, and behaviors [1]. CBT can address the cognitive distortions (CDs) that are related to biased, inaccurate, or irrational thoughts that have a significant influence on mental

Alian, M., Al-Khazaleh, M., AlShboul, H. (2025). Detection of Cognitive Distortions in Students' Thoughts Using Topic Modeling and Fuzzy Clustering. *International Journal of Interactive Mobile Technologies (iJIM)*, 19(16), pp. 22–40. https://doi.org/10.3991/ijim.v19i16.56253

Article submitted 2025-04-26. Revision uploaded 2025-06-25. Final acceptance 2025-06-26.

health and emotional regulation [2]. Distorted thoughts can allow negative reality perception that may eventually develop into mental health problems, such as depression and anxiety, according to the cognitive behavioral model [3].

In psychology, CDs are identified in patterns, such as all-or-nothing thinking (black-and-white thinking), mental filtering, emotional reasoning, disqualifying the positive, mind reading, self-blame, labeling, catastrophizing, and overgeneralization [4] [5].

After taking exams, students often experience various forms of distorted thinking patterns that exacerbate negative feelings about themselves. These patterns may include viewing things in extremes (all-or-nothing thinking), anticipating the worst possible outcomes (catastrophizing), focusing only on the negative and ignoring the positive (tunnel vision), making negative predictions about future outcomes without evidence (fortune-telling), or assuming that all subsequent future events will have the same outcome according to a negative event (overgeneralization).

These thinking patterns can amplify self-doubt and negative feelings during the exam process [6]. The core applications of machine learning (ML) in detecting and predicting potential mental health risks or distorted thoughts represent a significant advancement in mental health assessment and treatment. Natural language processing (NLP) is an essential field of ML that focuses on understanding, interpreting, and generating textual data in a useful way [7].

Clustering is a technique in supervised ML used to identify similarities in data and group them [8]. This technique is one of the essential tasks that can analyze a large amount of textual data collected from social media, group discussions, online forums, mental health questionnaires, therapy sessions, patient self-reports, clinical assessments, and treatment outcomes.

Furthermore, clustering can be helpful in CD analysis because it groups similar distorted thought patterns, recognizes and detects basic patterns in distorted thinking styles, identifies common distortion groups, and discovers new distorted groups in large datasets.

Topic modeling and fuzzy clustering are two commonly used clustering methods. Fuzzy clustering, also known as fuzzy C-means (FCM), is a type of clustering technique that provides a flexible approach to data grouping, where it is suitable for complex datasets that contain overlapping or uncertain data points; it allows a single data point to simultaneously encompass multiple groups with different weights relevant to each group [9]. CDs often interweave multiple types of distorted thinking patterns, and the borders between these types are not always clear. Fuzzy clustering emerges as a valuable analytical tool. This technique can recognize the complex nature of distorted thoughts by allowing each thought to be simultaneously associated with multiple distortion categories, each with varying weights. Accordingly, this technique can help in analyzing overlapped distorted thoughts.

Topic modeling, also known as topic analysis, is a critical technique used in different NLP applications; it excels at categorizing themes and patterns within a corpus of text and reveals the hidden structures within a text [10] [50]. When topic modeling is applied to CD analysis, it serves as a powerful tool for identifying various types of CD embedded in people's thoughts, such as overgeneralization, mental filtering, and all-or-nothing thinking.

A robust framework can be developed for an extensive understanding, identification, and management of CD patterns by combining CBT principles and advanced ML methods. Moreover, this framework provides valuable insights for mental health professionals and researchers, enabling the development of tailored and effective mental health care solutions.

Research on CD detection using ML techniques has demonstrated significant potential in various NLP applications. However, few studies have focused on analyzing Arabic text to detect CDs, presenting a valuable opportunity for investigation and development in this field.

This study examines how CDs manifest in students' thought patterns after exams. The common patterns of distorted thinking that arise in academic evaluation contexts are identified by applying topic modeling techniques and fuzzy clustering to students' self-reported reflections and thoughts. The methodology combines NLP techniques and pretrained embedding models to extract thematic elements from student narratives with fuzzy clustering that allows sophisticated categorization of overlapping CD types. Unlike classification methods, fuzzy clustering acknowledges that CDs present as blended phenomena.

The remainder of this study is as follows: Section 2 presents related work in the area of this study; Section 3 provides the definitions for the main concepts as a background; Section 4 describes the study methodology; Section 5 discusses the experiments and results; and Section 6 outlines the final conclusion.

2 RELATED WORK

CDs have been conceptualized as irrational thinking patterns that contribute to various psychological disorders [11]. This concept can be defined as a systematic error in perceiving information to convince a person of incorrect information [12] [13]. Researchers have increasingly explored automated methods to detect and classify these distorted thinking patterns due to the advancement of ML technologies. This section explores how researchers applied ML techniques to analyze CDs. We aim to identify knowledge gaps and understand the current state of the field by examining existing research. Our investigation focuses on several key aspects, including ML applications in CD analysis and psychological research, while highlighting the effective methodological approaches in this domain.

Recent studies have predominantly explored the use of ML and NLP techniques to identify CDs in text. The researchers have utilized various data sources to conduct these studies, such as personal blogs, journaling text, group discussions, online forums, mental health questionnaires, and therapy sessions. These studies utilized different feature extraction techniques and ML algorithms. Simms et al. [14] detected CD in personal blog posts from Tumblr API (493 posts) and utilized LIWC software for feature extraction. Various ML models were applied, and logistic regression (LR) had the best results with an accuracy of 73.0%.

Different studies utilized deep learning techniques, such as support vector machine (SVM) and convolutional neural network (CNN). Mostafa et al. [15] investigated automated methods of CD detection and classification in journaling texts. The proposed methodology utilized two techniques for feature extraction, including term frequency–inverse document frequency (Tf–Idf) and count vectorizers. This study investigated different ML models, including LR, SVM, and Naive Bayes, alongside deep learning models, such as CNN and LSTM. The results showed that the combination of LR with a count vectorizer achieved the highest performance among the ML models, with a 95% F1 score. However, the LSTM deep learning model with GloVe300d embeddings surpassed this to reach a 97% F1 score.

Shickel et al. [16] introduced an ML framework to automatically detect and classify 15 different types of CDs. This study utilized two mental health datasets: one obtained through crowdsourcing and the other from a real-world online therapy program. In the feature extraction, the term frequency-inverse document frequency (Tf–Idf) technique was utilized. The framework encompassed various ML models (LR, SVM, random forests, and gradient boosted trees) and two deep learning models (recurrent neural networks [RNN] and CNN). Additionally, exploratory analysis was conducted using two unsupervised learning models: an exploratory model and latent Dirichlet allocation (LDA). The authors found that LR achieved the best F1 score of 0.88 among all experiments.

Saba et al. [19] developed a hybrid model combining CNN and SVM for detecting and predicting depression-related mental illness through Arabic speech analysis. The comparative experiments were conducted using standalone RNN and CNN models alongside the proposed hybrid model. The results demonstrated the hybrid model's superiority, achieving an accuracy of 91.60%. Additionally, the hybrid model exhibited enhanced performance metrics, including reduced false positive and false negative rates and improved AUC, sensitivity, and specificity values compared with the models separately implemented.

Some studies used large language models (LLMs). For example, in addressing academic stress among undergraduate students, Indumini et al. [22] developed an AI chatbot to underscore the need for accessible and effective mental health needs. The study utilized OpenAI's GPT-3.5 turbo and various ML models, including SVM, multinomial LR, and random forest, to classify CDs, with the SVM model achieving the highest accuracy of 85.94%.

In addition, Ding et al. [17] analyzed therapy text messages from a community focusing on individuals with severe mental health illness. In this study, two approaches were utilized, namely, data augmentation techniques (specifically back translation and mix-up methods) and the MentalBERT model, to enhance the detection of uncommon CDs. The results showed that these techniques effectively worked for detecting uncommon CD classes, and the MentalBERT model demonstrated superior performance when identifying common CD classes.

Lim et al. [18] introduced a framework named ERD to improve the performance of LLMs in classifying CDs by extracting relevant parts of the text and multi-agent debating reasoning steps. A public dataset from Kaggle was used for the evaluation process, which contained 2530 instances. The results showed that ERD improved the binary specificity score and multi-class F1 score for the LLM-based CD classification, especially by reducing the high false positive rate observed with baseline techniques.

Wang et al. [20] conducted experiments using the first Chinese language dataset of CDs, consisting of 7500 sentences collected by individuals with psychological training and validated and classified by psychology experts. This study compared the performance of various models, including ChatGPT and deep learning models, such as BERT and RoBERT. The RoBERT model showed superior performance in CD classification, achieving an F1-score of 0.73. Meanwhile, ChatGPT demonstrated relatively lower performance. The incorporation of CD features improved the performance of mental disorder classification models. This improvement increased the F1-scores in detecting depression (from 0.73 to 0.79) and post-traumatic stress disorder (from 0.74 to 0.85).

Lin et al. [21] developed a positive reconstruction framework along with a Mandarin Chinese dataset designed for detecting CDs and generating positively reframed alternatives. The experiments utilized 4001 samples for CD detection and 1900 samples for the positive reframing task. Various LLMs were used to evaluate the implementation of fine-tuning and prompt engineering techniques. The findings revealed that fine-tuning these models produced more effective results than prompt engineering. Among all the tested models tested, the fine-tuned version of ChatGPT-3.5 demonstrated superior performance, achieving the highest BLEU score of 10.68 on the positive reconstruction task.

Table 1 provides a comparison between the previously discussed research in terms of year of publication, applied models, extracted features, dataset, evaluation metrics, and results.

Table 1. Comparison between related works

Ref.	Year	Models Utilized	Feature Extraction	Dataset	Evaluation Metrics	Best Results
[14]	2017	Decision tree, logistic regression, naïve Bayes, and k-nearest neighbors	Correlation-based feature subset selection (CFS) and RELIEF	Personal blog posts from Tumblr API	Accuracy	Logistic regression 73.0%
[15]	2021	– ML models (LR, SVM, and NB) – Deep learning models (CNN and LSTM)	Tf-idf vectorizer and count vectorizer	The dataset was collected from (Twitter, a survey) and the HappyDB dataset	F1 Precision Recall	A fine-tuned LSTM with GloVe300d embeddings achieved a 97% F1 score
[16]	2020	– ML models (LR, SVM, RF, GBoost) – Deep learning models (RNN, CNN) – Unsupervised learning models (exploratory and LDA)	Tf-IDF	Two mental health datasets, one collected from crowdsourcing and another from a real-world online therapy program	F1 Precision Recall	Logistic regression achieved a 0.88 F1 score
[17]	2022	Data augmentation techniques (back translation and mixup methods) and the MentalBERT	N/A	Public dataset from Kaggle	– Area under the precision-recall curve (AUPRC)	Mixup detecting uncommon cognitive distortions classes (improved AUPRC results by 1.6%)
[18]	2024	LLM (GPT-3.5-turbo)	LLM-based identification of specific relevant text segments acts as an intelligent filter	Dataset from Kaggle (2530 instances)	F1 Specificity Sensitivity	ERD without summarization achieved a 92.13 sensitivity score and a 75.48 F1 score
[19]	2022	CNN and SVM	Spectral centroid and spectral contrast features	Basic Arabic Vocal Emotions Dataset (BAVED)	Accuracy	RNN and CNN models alongside the hybrid model achieved an accuracy of 91.60%
[20]	2023	ChatGPT BERT RoBERT	C2D2	Chinese language dataset of cognitive distortions	F1	RoBERT achieved F1 score of 0.73
[21]	2024	LLMs	Implicitly handled by the RoBERTa pre-trained language model	Mandarin Chinese dataset	BLEU	ChatGPT-3.5 achieved a BLEU score of 10.68
[22]	2024	GPT-3.5 Turbo, SVM, Multinomial Logistic Regression, and Random Forest	Labeled words are analyzed by SVM and transformed to high-dimensional feature vectors	Intent classification dataset and cognitive distortion identification dataset	Accuracy	The SVM model achieved the highest accuracy of 85.94%

This study aims to identify CDs in the thinking patterns of university students after completing their exams. The methodology combines topic modeling with fuzzy clustering techniques, utilizing AraBERT and AraGPT models to convert Arabic text into embeddings for analysis.

3 BACKGROUND

3.1 Clustering

The term "clustering" refers to a technique that divides unlabeled data into distinct groups with little to no supervision. The result of the grouping process indicated that the items in the same cluster differ from those in other clusters while sharing comparable properties [23] [51].

Clustering has also been defined as an ML component that addresses unsupervised learning. The learning process is based on algorithms that identify patterns in the dataset that are either simulated or derived from an actual observation [24] [25]. In this unsupervised learning technique, the data are divided into groups based on some similar characteristics and features in the grouped data. This technique is used in several NLP applications, such as sentiment analysis, paraphrasing identification, and question answering [26].

Furthermore, this technique can be helpful in detecting CDs by identifying types of disordered thoughts or patterns based on specific features because it is considered a powerful tool in textual data analysis by grouping unlabeled data to extract valuable information [27].

Topic modeling. Topic modeling is considered the latest development in text mining. This statistical technique is used for identifying the underlying semantic structure in a sizable document collection [28]. Large textual datasets can be automatically organized, interpreted, and summarized using topic modeling and information retrieval techniques, which uncover latent topics within a collection of documents [29].

In different NLP applications, topic modeling provides an interpretable presentation of documents, where two different approaches are used with topic modeling: LDA and non-negative matrix factorization (NMF) [29].

However, LDA is considered the most often used topic modeling technique [52]. LDA is a generative probabilistic model where each item is presented as a finite mixture over an underlying set of topics, and each topic is presented as an infinite mixture over a collection of topic probabilities [30]. Moreover, LDA is a three-level hierarchical Bayesian model that provides an effective tool to generate an explicit representation for a document because the number of topics does not have to be predetermined [31]. Meanwhile, NMF is an unsupervised technique, where the subjects that the model will be trained on are not labeled [32]. This decomposition, a non-probabilistic technique, uses matrix factorization, where non-negative values are stored [31].

Fuzzy clustering. The clustering technique, known as "fuzzy clustering," is based on Zadeh's fuzzy set [33], and it is one of the unsupervised techniques where a single point of data can belong to more than one group at the same time; accordingly, each point has a member degree to each cluster [23]. This technique is widely used in analyzing complex data, such as psychological data [34].

The aforementioned technique is sometimes referred to as soft clustering, which is an incredibly practical technique and has various uses, such as recommending movies and grouping consumers according to their interests. However, consumers may be interested in multiple genres. If only one sort of content is suggested, then they may feel irritated. Therefore, fuzzy clustering is appropriate in this case, as it groups data in a number of categories to indicate how much a data object belongs to a particular class; it is given a membership [35] [36].

3.2 Evaluation metrics

Different metrics can be used to evaluate the results of topic modeling, such as silhouette score, topic coherence, topic diversity, and intra- and inter-clustering. However, topic coherence and topic diversity are commonly used to evaluate the resulting clusters from topic modeling. Additional details about these metrics are provided in the next subsections.

Topic coherence. Coherence score is one of the evaluation metrics that can be used to show how well topic modeling performs. This metric is typically utilized to examine the resemblance or link between two datasets. Topic coherence in topic modeling assesses the quality of the data by analyzing the semantic similarity of highly repetitious words in a topic [37]. The coherence score is a number between zero and one, where zero represents poor coherence (low similarity), and one denotes good coherence (high similarity). Additionally, two datasets that are identical and related will have strong coherence, while datasets without any association are considered to have poor coherence [38].

Silhouette score. Silhouette score is a technique for interpreting and verifying consistency among data clusters. This technique was proposed in 1987 by Peter Rousseeuw [39], a statistician from Belgium. The technique offers a concise visual depiction of each object's classification accuracy [39]. This popular metric is used to evaluate the quality of clustering by measuring the distance between clusters and cohesion within them [40]. Although the silhouette score is a useful heuristic, it has certain limitations, particularly when applied to noisy or high-dimensional data. Furthermore, the score reliability of this metric decreases when clusters overlap or have irregular sizes [40].

Topic diversity. Topic diversity was introduced to quantify the differences between topics. This metric is motivated by the expectation that subjects should be diverse rather than redundant to fully reveal hidden semantics in the corpora [41].

Diversity metrics use the uniqueness of individual words to measure topic diversity. Several studies [42] [43] [41] have indicated that diversity is at its best when each topic is distinguished by a unique top word.

Intra- and inter-cluster similarity. The assessment of clustering quality is known as cluster validity. In a high-quality cluster, the intra-cluster distance is maintained at a minimum value, and the inter-cluster distance is kept at a maximum value [44].

Intra-cluster distance describes the typical separation between data points that constitute the same cluster. Additionally, this intra-cluster distance evaluates how cohesive or compact a cluster of data points is. The more similar and nearby the data points within a cluster, the lower the intra-cluster distance, where the average or maximum distance between pairs of data points within a cluster is commonly used to compute intra-cluster distance [45] [46].

Inter-cluster distance describes the typical separation between several clusters within resulted in a clustering solution. This mechanism calculates how different or distant clusters are from one another. Clusters are considered to be clearly defined and well-separated when the inter-cluster distance is high. Moreover, the inter-cluster distance is typically computed as the minimum distance between data points in distinct clusters or as the distance between the centroids of the clusters [45] [46].

4 METHODOLOGY

Two models of topic modeling are used to determine CDs in students' thoughts after taking exams, namely, LDA with TF–Idf vectorizer and NMF with two embeddings (AraBERT and AraGPT). Thereafter, the models are combined with fuzzy clustering. However, LDA is a probabilistic topic modeling that works with the word–document frequency matrix that Tf–Idf provides. Meanwhile, AraBERT and AraGPT generate numerical word vectors rather than frequency-based ones. Therefore, NMF is used because it effectively works with deep-learning embeddings.

LDA topic modeling, as illustrated in Figure 1, starts by loading a dataset, followed by data preprocessing, including stop word removal, removing punctuation, and eliminating non-Arabic words, personal names, numerical values, diacritical marks, and special characters. Additionally, tokenization is performed, and normalization is applied for repeated letters and letters with different forms to standardize them and ensure data quality.

Thereafter, text vectorization is performed using Tf–Idf. After the vectorization, topic modeling is performed using LDA. Topics with top words are displayed and assigned to each sentence. Afterward, fuzzy clustering is applied using the C-means algorithm. The sentences are assigned to the cluster with the highest membership probability to evaluate the clustering results using coherence score and other metrics.

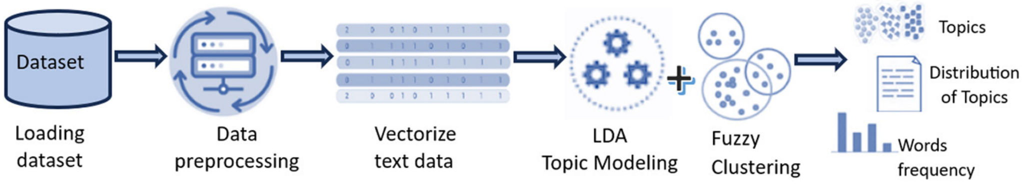


Fig. 1. LDA topic modeling with fuzzy clustering

Topic modeling with NMF and a pretrained embedding model starts by loading a dataset. Thereafter, a tokenizer that is suitable for embedding a model is loaded as shown in Figure 2. Subsequently, the pre-trained embedding model (AraBERT or AraGPT) is loaded. Thereafter, sentence embeddings are generated. We performed rectified linear unit (ReLU) transformation for the AraBERT/AraGPT embedding to ensure non-negative values for applying NMF in topic modeling. Afterward, FCM clustering is applied.

ReLU ensures that the input to NMF is sparse and non-negative, and it simplifies feature extraction before NMF reduces dimensions. Fuzzy clustering refines soft assignments, thereby ensuring scalability. These mechanisms work together to create a coherent, efficient, and interpretable framework that is ideal for addressing complicated, high-dimensional data.

Each sentence is assigned to the cluster with the highest membership value to evaluate the clustering quality using evaluation metrics, such as coherence, inter- and intra-clustering similarity, and other metrics.

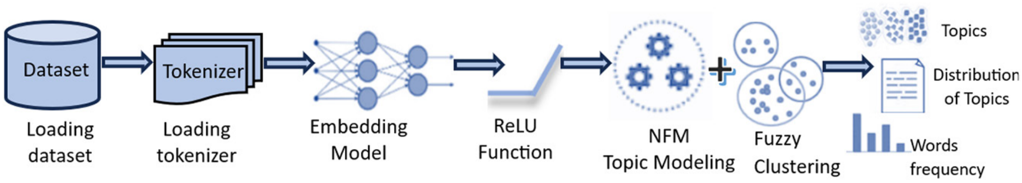


Fig. 2. NFM topic modeling with fuzzy clustering

5 EXPERIMENT AND RESULTS

The experiments in this study are conducted on two CD datasets by applying topic modeling based on different approaches, namely, Tf–Idf with LDA and NMF with two pre-trained embeddings, including AraBERT and AraGPT. Thereafter, the experiments are repeated using fuzzy clustering with topic modeling.

The results of all experiments are evaluated using five metrics: coherence score, average intra-cluster similarity, inter-cluster similarity, topic diversity, and silhouette score. The traditional topic modeling is compared using these approaches.

The two datasets utilized to conduct our experiments reflect the common CDs among students after exams. The first one is a students' dataset that consists of students' opinions after exams, while the other is a generated dataset based on the principles of CBT as demonstrated in different books.

5.1 Datasets

Student dataset. The students' dataset is collected from responses by students who expressed emotions and feelings after taking academic exams. This dataset is particularly useful for CD prediction, which aligns with the objective of our study in detecting distorted thinking patterns. The dataset contains 200 responses or opinions after taking an exam. The instances with missing or irrelevant information are removed, such as empty responses and responses where only students' names are provided without emotional content. Table 2 presents examples of cognitive thinking patterns observed from the student's responses.

Table 2. Examples of distorted thinking patterns

Pattern	Example	
All-or-nothing thinking	Either I succeeded with distinction, or I don't deserve success.	اما ان انجح بتفوق او انني لا استحق النجاح.
Catastrophizing	Failing this exam means that my life is over.	رسوبي في هذا الامتحان يعني ان حياتي انتهت.
Tunnel vision	Why should I try if I am going to fail anyway?	لماذا احاول سافشل على اي حال.
Fortune-telling	Everyone is sure I failed.	الجميع متاكد انني رسبت.
Overgeneralization	Any small mistake means that I won't get the result that I want.	اي خطا صغير يعني انني لن احصل على النتيجة التي اريدها.

Generated dataset. The generated dataset is built based on the principles of CBT as defined and demonstrated in the work of Burns [47] and Beck [48] that explain how to identify and correct cognitive abnormalities in psychotherapy. This dataset consists of 300 sentences reflecting common CDs among students after exams and is manually categorized according to known criteria for each type of distortion.

Type–token ratio. Type–token ratio (TTR) [49] is used as a metric to measure lexical diversity in the responses of the students to analyze both datasets. TTR is measured as the ratio between the number of unique tokens and the total number of tokens. This ratio reflects how varied or repetitive the vocabulary in the dataset. Table 3 shows the number of all tokens, unique tokens, and the TTR ratio in each dataset.

Table 3. TTR ratio in both datasets

Dataset	Total Tokens	Unique Tokens	TTR Ratio
Student Dataset	4510	1725	0.3825
Generated Dataset	1797	125	0.0696

The TTR ratio for the students' thoughts dataset (0.38) is a moderate ratio, which represents that the students repeat numerous words when they write their thoughts. Additionally, the ratio indicates that students expressed similar CDs. Although the dataset exhibits a certain diversity, the CDs are dominant, reflecting different negative emotions, such as stress and self-doubt.

However, the TTR for the generated dataset is very low (0.07), indicating the presence of frequently used words in the sentences. This notion suggests that the dataset may contain similar patterns of cognitive thoughts.

5.2 Results and discussion

After applying the traditional topic modeling to the dataset of student's thoughts, the results in Figure 3 show that the utilization of NMF with AraBERT embeddings achieves the best coherence score (0.78). This notion means that the words inside each topic are semantically connected and related to each other. In addition, the utilization of NMF with AraGPT embedding shows a strong coherence (0.72), while the use of TF–Idf with LDA results in poor coherence (0.15), indicating that the topics lack meaningfulness.

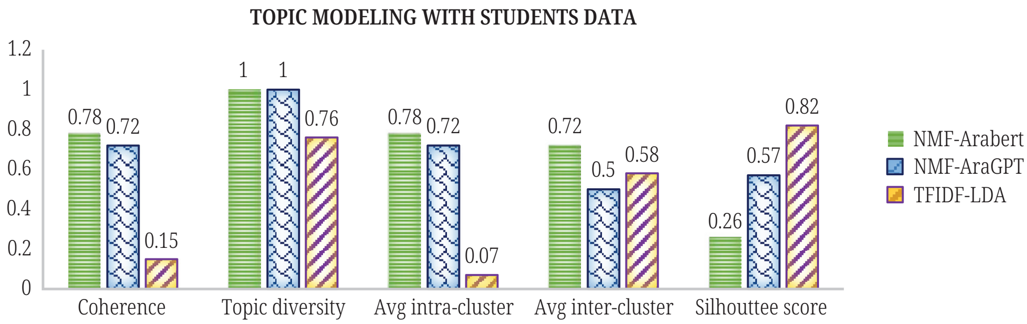


Fig. 3. Topic modeling results for student's thoughts dataset

The results in terms of the diversity metric of topic modeling based on NMF with embeddings AraBERT and AraGPT show a perfect score value of (1.0), which means no overlapping between topics. Meanwhile, topic modeling with Tf–Idf and LDA has a diversity score of 0.76, which means that some overlap exists between topics.

Figure 3 demonstrates that the topic modeling approach using NMF and AraBERT provides the highest intra-cluster similarity (0.78), indicating that the vectors representing sentences in the cluster are highly related. The results of using NMF and AraGPT have a lower value of intra-cluster similarity (0.72) compared with NMF–AraBERT. Meanwhile, the use of TF–Idf with LDA has the worst value (0.07), indicating that the sentences in the cluster (i.e., topic) are different.

The NMF with AraGPT achieves the lowest inter-cluster similarity score of 0.5, indicating that the clusters (topics) are well separated, indicating optimal performance. The use of NMF and AraBERT achieves a moderate value of inter-cluster similarity (0.72), indicating that the topics are moderately distinct. Meanwhile, using TF–Idf with LDA provides a score of (0.58), denoting that some overlap exists between topics.

In terms of silhouette score, using TF–IdF with LDA achieves the best score (0.82), while using NMF with AraGPT provides a moderate score, and NMF with AraBERT has the lowest score. However, given that the silhouette score effectively works with approaches that use distance, it has limitations with approaches that utilize similarity.

The results of applying topic modeling to the generated dataset are illustrated in Figure 4. The NMF with AraBERT provides the highest coherence score (0.86), which means that the generated topics are semantically meaningful and related to each other. Additionally, NMF with AraGPT shows a coherence score of 0.70, which is lower than that of NMF–AraBERT but is still comparatively strong. Meanwhile, LDA shows a very low coherence score (0.06), which means that it fails to generate meaningful topics.

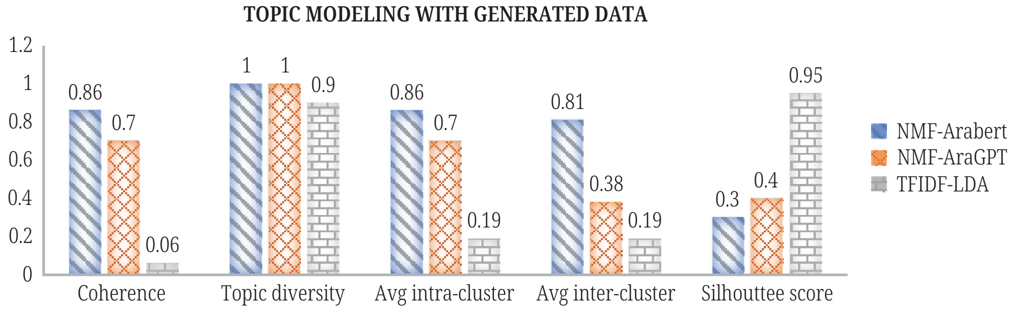


Fig. 4. Topic modeling results with a generated dataset

In terms of topic diversity, NMF with both types of embeddings, namely, AraBERT and AraGPT, shows the maximum topic diversity value, which indicates that the topics are not repetitive and have a wide range of words. Moreover, the LDA shows a strong topic diversity value of 0.9.

Figure 4 shows that NMF with AraBERT has a silhouette score of 0.31, which indicates a moderate topic separation. Meanwhile, NMF with AraGPT provides an improved silhouette score (0.40) compared with NMF–AraBERT, which means improved defined topic clusters. Meanwhile, LDA achieves the highest silhouette score (0.95), indicating a well-separated topic.

The highest intra-cluster similarity is achieved by NMF with AraBERT (0.86), which indicates that the sentences clustered in a topic are similar, while NMF with AraGPT achieves a lower value (0.70), denoting less cohesion inside topics. LDA with Tf–Idf provides a poor intra-cluster similarity (0.19), which means that the sentences within a topic are not well-grouped.

The scores of inter-cluster similarity show a high value achieved by NMF with AraBERT (0.81), which indicates less topic separation and topic overlapping. However, NMF with AraGPT achieves a much lower score (0.38), while LDA with TF–Idf has the lowest inter-cluster similarity score (0.19), which means a clear separation of topics for both.

Figure 5 presents the results of applying fuzzy clustering with two topic modeling approaches, namely, NMF with two embeddings (AraBERT and AraGPT) and LDA with TF–Idf, to identify the students' cognitive thoughts after taking exams. NMF with AraBERT achieves a perfect coherence score, indicating that the topics are meaningful and well-structured; nonetheless, it has a lower topic diversity score (0.58). Moreover, this approach provides high topic cohesion because of the high intra-cluster similarity score (0.81), but it also exhibits topic overlap, reflected in the high inter-topic similarity score (0.74). NMF–AraBERT provides a silhouette score of 0.35, indicating a moderate clustering quality but not well separated.

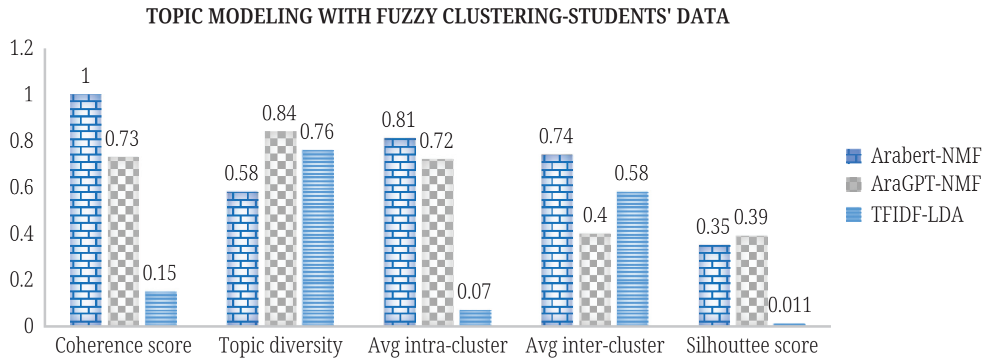


Fig. 5. Topic modeling with fuzzy clustering results

NMF with AraGPT achieves a coherence score of 0.73, which is lower than the score achieved by MMF with AraBERT but still high, while it has higher topic diversity (0.84), indicating varied topics as shown in Figure 5. Additionally, this approach provides lower scores for intra-cluster similarity (0.72) and inter-cluster similarity than those achieved by NMF with AraBERT, indicating less cohesion but better separated topics. Using AraGPT provides a higher silhouette score (0.39) than AraBERT, indicating that the generated clusters are well defined.

However, LDA with Tf–Idf and fuzzy clustering shows a very low intra-clustering similarity (0.07) but a moderate inter-clustering similarity score (0.58), indicating poor cohesion and moderate overlap within topics. Moreover, this approach has a very poor silhouette score (0.01), but it provides a topic diversity value of 0.76, which is lower than NMF with AraGPT but higher than NMF with AraBERT.

In the generated dataset, the NMF with AraBERT generates highly interpretable topics, as shown by its coherence score (1), but it has a lower diversity score than other models. This approach achieves high scores for intra-cluster and inter-cluster similarity (0.86 and 0.83, respectively), indicating that the topics contain very similar sentences and have similarity with other topics, which means overlapping topics. Moreover, NMF–AraBERT has a moderate clustering quality, as shown by the silhouette score (0.43) as shown in Figure 6.

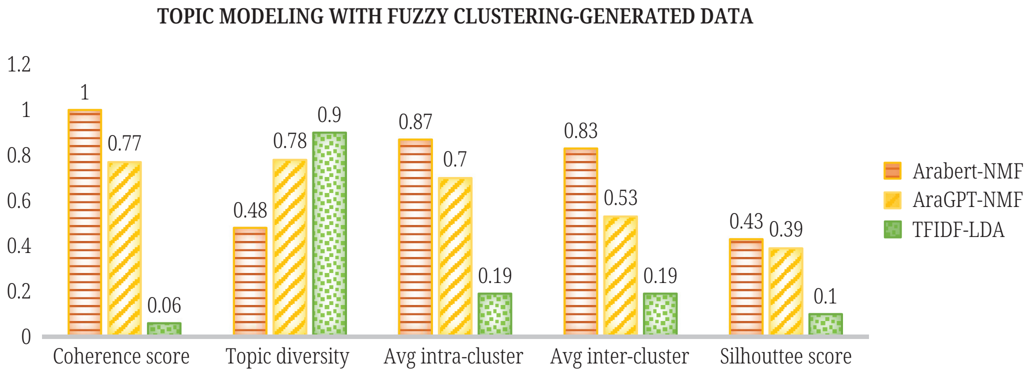
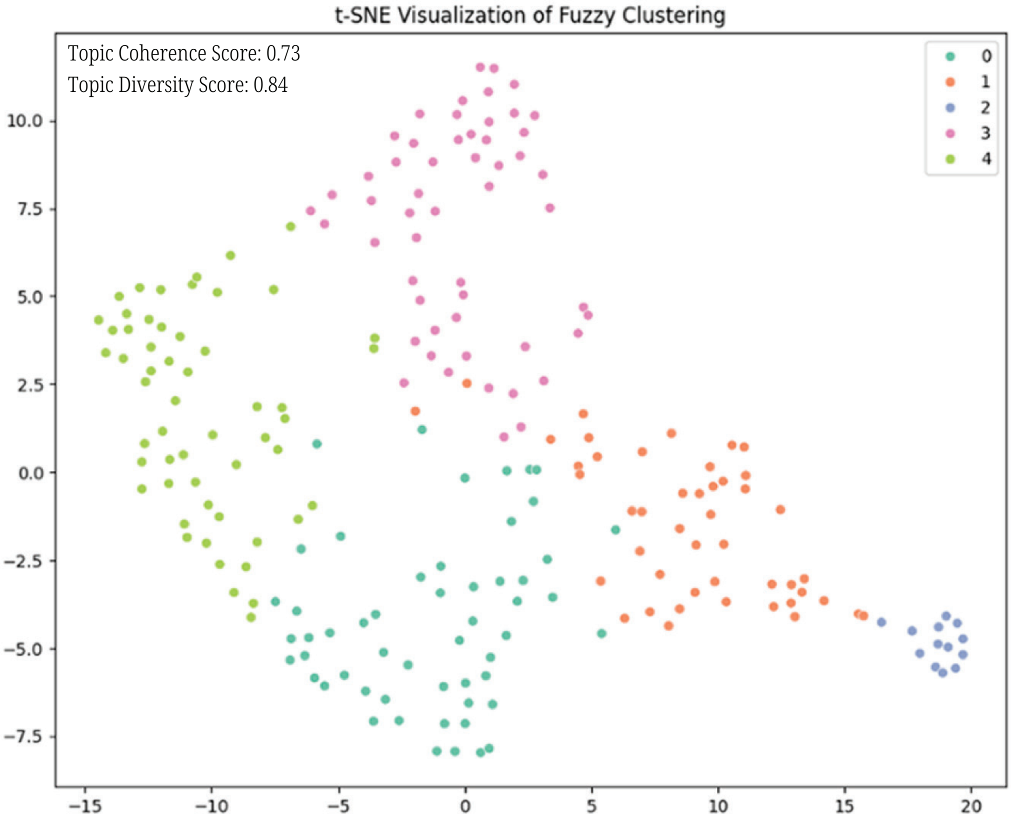


Fig. 6. Results of topic modeling with fuzzy clustering for the generated dataset

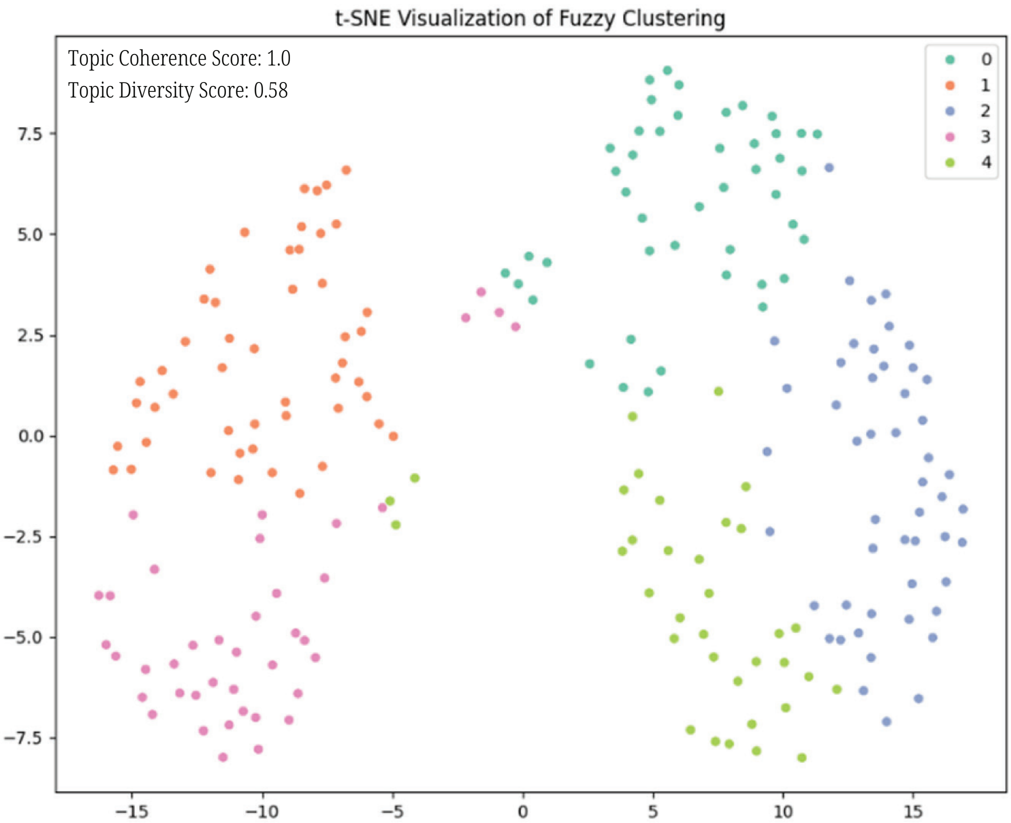
The NMF with the AraGPT model has a lower coherence score (0.7714) than AraBERT with NMF, but it has a higher topic diversity score (0.78), which means that it generates more distinct topics, as shown in Figure 6. Additionally, this approach provides less intra-cluster similarity (0.70) than AraBERT with NMF, which indicates less cohesion within topics; however, it has better inter-cluster similarity (0.53), which means better separation between topics than AraBERT with NMF. Furthermore, this approach has a moderate silhouette score of 0.39.

The results of fuzzy clustering combined with Tf–Idf–LDA show an extremely low coherence score (0.06), which indicates that the topic lacks significance. This approach has a high diversity score (0.9) but very low scores for intra-cluster and inter-cluster similarity, which means poor topic cohesion and low topic overlapping. In addition, this approach achieves a very poor silhouette score (0.096), indicating that the topics are randomly assigned.

Figures 7 and 8 visualize the resulting clusters from topic modeling combined with fuzzy clustering, where the model NMF with two pre-trained embeddings, namely, AraBERT and AraGPT for students' datasets and the generated dataset, respectively, is utilized.

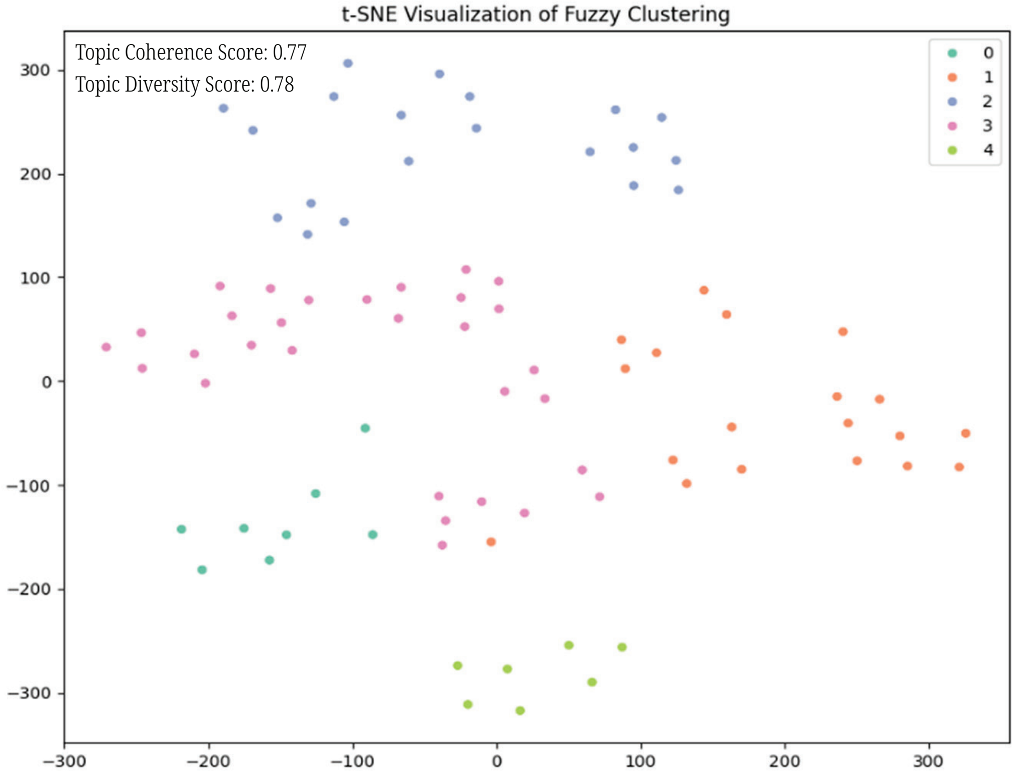


a) Students dataset NMF-AraGPT

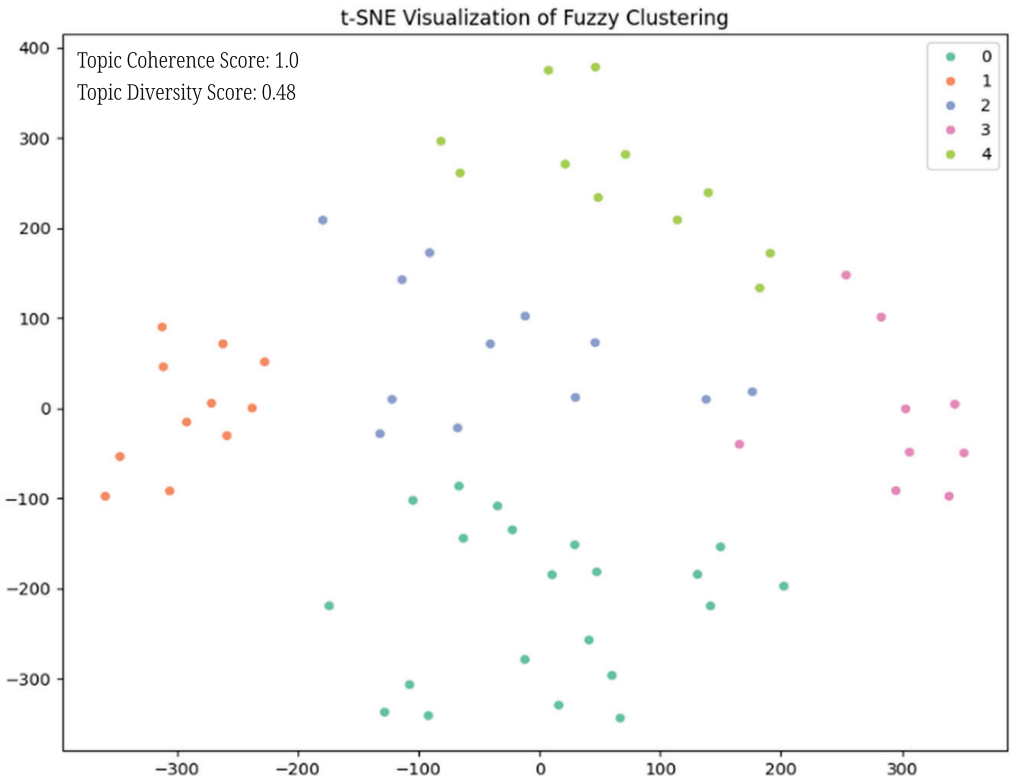


b) Students dataset NMF-Arabert

Fig. 7. Visualization of the results of fuzzy clustering with topic modeling for the student dataset



a) Generated dataset NMF-AraGPT



b) Generated dataset NMF-Arabert

Fig. 8. Visualization of the results of fuzzy clustering with topic modeling for the generated dataset

The t-SNE visualization is used to visualize the results of fuzzy clustering, where t-SNE is an effective dimensionality reduction algorithm used to visualize high-dimensional data in a low-dimensional space. Topic modeling with NMF with AraGPT is the most effective model when combined with fuzzy clustering.

This study utilized fuzzy clustering as an unsupervised topic modeling approach to identify CDs in students' written thoughts, where no ground truth is available. Unlike previous studies mentioned in the related work section, which utilized supervised ML approaches, which depend on annotated datasets and used F1-score, precision, and accuracy as evaluation metrics. However, this study aimed to discover latent topic structures without prior labeling.

The fuzzy clustering approach provides great flexibility and interpretability, especially in exploratory settings or when labeled data are limited or unavailable. Moreover, this approach allows soft membership, where each document or sentence can belong to multiple topics (i.e., clusters) to varying degrees. This approach provides an advantage when dealing with psychological constructs, such as CDs. In terms of scalability, fuzzy clustering is easy to adapt to new data, while previous studies require new annotations.

However, we acknowledge that unsupervised evaluation metrics, such as coherence or silhouette score, may not directly correspond to real-world classification accuracy. Unsupervised and supervised approaches are not inherently better or worse; instead, they provide complementary perspectives. The selection of an appropriate method depends on certain factors, such as the availability of annotated data, the goal of the analysis, and the required interpretability.

6 CONCLUSION

This study focuses on the application of topic modeling with fuzzy clustering to detect CDs in students' thoughts after taking exams. The experiments show that the application of topic modeling with fuzzy clustering and NMF-AraGPT to students' datasets is an effective approach because it balances coherence and diversity while generating distinct and separated topics. Although using fuzzy clustering with NMF–AraBERT produces the most coherent topics (0.81), it has low diversity and high inter-topic similarity scores, indicating some redundancy. The generated topics are not well separated, as indicated by the silhouette score of 0.35. However, topic modeling using LDA with Tf–Idf exhibits poor performance with fuzzy clustering. This approach did not generate meaningful topics because of its low coherence score (0.15), and it achieved a poor clustering quality, as shown by its silhouette score of 0.01.

In the generated dataset, topic modeling with NMF and AraBERT generates coherent and well-structured topics. However, numerous topics overlap. Meanwhile, AraGPT with NMF improves the balance between coherence and diversity for the generated dataset. However, fuzzy clustering with Tf–Idf and LDA fails to produce meaningful topics for this dataset, as shown by its low coherence score (0.05), and provides poor scores for all metrics.

In future work, practical applications would be included to enhance the scope of topic modeling and fuzzy clustering, which can directly benefit educational institutions and students by developing systems that can detect CDs in real-time students communications.

7 REFERENCES

[1] A. Wenzel, "Strategies of cognitive behavioral therapy," *Psychiatric Clinics of North America,* vol. 40, no. 4, pp. 597–609, 2017. https://doi.org/10.1016/j.psc.2017.07.001

[2] E. E. Fortune and A. S. Goodie, "Cognitive distortions as a component and treatment focus," *Psychology of Addictive Behaviors*, vol. 26, no. 2, pp. 298–310, 2012. https://doi.org/10.1037/a0026422

[3] J. S. Beck, *Cognitive Behavior Therapy: Basics and Beyond*. New York, NY: Guilford Publications, 2020. https://psycnet.apa.org/record/2020-66930-000

[4] P. Gilbert, "The evolved basis and adaptive functions of cognitive distortions," *British Journal of Medical Psychology,* vol. 71, no. 4, pp. 447–463, 1998. https://doi.org/10.1111/j.2044-8341.1998.tb01002.x

[5] E. Kuru *et al.*, "Cognitive distortions in patients with social anxiety disorder: Comparison of a clinical group and healthy controls," *The European Journal of Psychiatry*, vol. 32, no. 2, pp. 97–104, 2018. https://doi.org/10.1016/j.ejpsy.2017.08.004

[6] S. H. Katherine, T. C. John, and J. S. Stuart, "Maladaptive perfectionism, impostorism, and cognitive distortions: Threats to the mental health of pre-clinical medical students," *Academic Psychiatry*, vol. 43, pp. 381–385, 2019. https://doi.org/10.1007/s40596-019-01031-z

[7] M. J. Al-Khazaleh, M. Alian, and M. Jaradat, "Sentiment analysis of imbalanced Arabic data using sampling techniques and classification algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 607–618, 2024. https://doi.org/10.11591/eei.v13i1.5886

[8] E. Feczko and D. A. Fair, "Methods and challenges for assessing heterogeneity," *Biological Psychiatry,* vol. 88, no. 1, pp. 9–17, 2020. https://doi.org/10.1016/j.biopsych.2020.02.015

[9] M.-C. Hung and D.-L. Yang, "An efficient Fuzzy C-Means clustering algorithm," in *Proceedings 2001 IEEE International Conference on Data Mining*, San Jose, CA, USA, 2001, pp. 225–232. https://doi.org/10.1109/ICDM.2001.989523

[10] A. Meddeb and L. Ben Romdhane, "Using topic modeling and word embedding for topic extraction in twitter," *Procedia Computer Science*, vol. 207, pp. 790–799, 2022. https://doi.org/10.1016/j.procs.2022.09.134

[11] C. L. Yurica and R. A. DiTomasso, "Cognitive distortions," in *Encyclopedia of Cognitive Behavior Therapy*, Boston, MA: Springer, 2005, pp. 117–122. https://doi.org/10.1007/0-306-48581-8_36

[12] B. Wang, P. Deng, and Y. Zhao, "C2D2 dataset: A resource for the cognitive distortion analysis and its impact on mental health," *Findings of the Association for Computational Linguistics*, pp. 10149–10160, 2023. https://doi.org/10.18653/v1/2023.findings-emnlp.680

[13] I. P. G. H. Suputra *et al.*, "Detection and classification of cognitive distortions: A literature review," in *International Conference on Smart-Green Technology in Electrical and Information Systems (ICSGTEIS)*, Badung, Bali, Indonesia, 2023, pp. 166–171. https://doi.org/10.1109/ICSGTEIS60500.2023.10424225

[14] T. Simms, C. Ramstedt, M. Rich, M. Richards, T. Martinez, and C. Giraud-Carrier, "Detecting cognitive distortions through machine learning text analytics," in *IEEE International Conference on Healthcare Informatics (ICHI)*, Park City, UT, USA, 2017, pp. 508–512. https://doi.org/10.1109/ICHI.2017.39

[15] M. Mostafa, A. El Bolock, and S. Abdennadher, "Automatic detection and classification of cognitive distortions in journaling text," in *Proceedings of the 17th International Conference on Web Information Systems and Technologies WEBIST*, vol. 1, 2021, pp. 444–452. https://doi.org/10.5220/0010713000003058

[16] B. Shickel, S. Siegel, M. Heesacker, S. Benton, and P. Rashidi, "Automatic detection and classification of cognitive distortions in mental health text," in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2020, pp. 275–280. https://doi.org/10.1109/BIBE50027.2020.00052

[17] X. Ding, "Improving classification of infrequent cognitive distortions: Domain-specific model vs. Data augmentation," in *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Student Research Workshop*, 2022, pp. 68–75. https://doi.org/10.18653/v1/2022.naacl-srw.9

[18] S. Lim, Y. Kim, C.-H. Choi, J.-Y. Sohn, and B.-H. Kim, "ERD: A framework for improving LLM reasoning for cognitive distortion classification," in *Proceedings of the 6th Clinical Natural Language Processing Workshop,* 2024, pp. 292–300. https://doi.org/10.18653/v1/2024.clinicalnlp-1.25

[19] T. Saba, A. R. Khan, I. Abunadi, S. A. Bahaj, H. Ali, and M. Alruwaythi, "Arabic speech analysis for classification and prediction of mental illness due to depression using deep learning," *Computational Intelligence and Neuroscience*, no. 1, 2022. https://doi.org/10.1155/2022/8622022

[20] B. Wang, P. Deng, Y. Zhao, and B. Qin, "C2D2 dataset: A resource for analyzing cognitive distortions and its," in *Conference on Empirical Methods in Natural Language Processing*, 2023, pp. 10149–10160. https://doi.org/10.18653/v1/2023.findings-emnlp.680

[21] S. Lin, Y. Wang, J. Dong, and S. Ni, "Detection and positive reconstruction of cognitive distortion sentences: Mandarin dataset and evaluation," in *Findings of the Association for Computational Linguistics*, Bangkok, Thailand, Association for Computational Linguistics, 2024. https://doi.org/10.18653/v1/2024.findings-acl.399

[22] M. M. A. H. Indumini, K. A. S. H. Kulathilake, and D. K. Hettiarachchi, "Implementing cognitive behavioral therapy in chatbots to reduce students' exam stress using ChatGPT," in *The SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI)*, Ratmalana, Sri Lanka, 2024, pp. 1–6. https://doi.org/10.1109/SLAAI-ICAI63667.2024.10844930

[23] A. E. Ezugwu *et al.*, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, p. 104743, 2022. https://doi.org/10.1016/j.engappai.2022.104743

[24] M. Ambigavathi and D. Sridharan, "Analysis of clustering algorithms in machine learning for healthcare data," in *Advances in Computing and Data Sciences*, M. Singh, P. Gupta, V. Tyagi, J. Flusser, T. Ören, and G. Valentino, Eds., Springer, vol. 1244, 2020, pp. 117–128. https://doi.org/10.1007/978-981-15-6634-9_12

[25] C. Cotton *et al.*, "An overview of clustering methods with guidelines for application in mental health research," *Psychiatry Research,* p. 115265, 2023. https://doi.org/10.1016/j.psychres.2023.115265

[26] M. Alian and R. H. AlMahmoud, "The effect of clustering algorithms on question answering," *Expert Systems with Applications,* vol. 243, p. 112959, 2024. https://doi.org/10.1016/j.eswa.2023.122959

[27] B. Shickel, S. Siegel, M. Heesacker, S. Benton, and P. Rashidi, "Automatic detection and classification of cognitive," in *IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2020, pp. 275–280. https://doi.org/10.1109/BIBE50027.2020.00052

[28] P. Kherwa and P. Bansal, "Topic modeling: A comprehensive review," *EAI Endorsed Transactions on Scalable Information Systems*, vol. 20, no. 24, p. e2, 2018. https://doi.org/10.4108/eai.13-7-2018.159623

[29] A. Abdelrazek, Y. Eid, E. Gawish, W. Medhat, and A. H. Yousef, "Topic modeling algorithms and applications: A survey," *Information Systems,* vol. 112, p. 102131, 2022. https://doi.org/10.1016/j.is.2022.102131

[30] S. J. Blair, Y. Bi, and M. D. Mulvenna, "Aggregated topic models for increasing social media topic coherence," *Appl. Intell.,* vol. 50, no. 1, pp. 138–156, 2020. https://doi.org/10.1007/s10489-019-01438-z

[31] R. Egger and J. A. Yu, "Topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts," *Front. Sociol.,* vol. 7, 2022. https://doi.org/10.3389/fsoc.2022.886498

[32] S. Mifrah and E. H. Benlahmar, "Topic modeling coherence: A comparative study between LDA and NMF models using COVID'19 corpus," *International Journal of Advanced Trends in Computer Science and Engineering,* vol. 9, no. 4, 2020. https://doi.org/10.30534/ijatcse/2020/231942020

[33] L. Zadeh, "Fuzzy sets," *Information and Control,* vol. 8, no. 3, pp. 338–353, 1965. https://doi.org/10.1016/S0019-9958(65)90241-X

[34] H. Han, "Fuzzy clustering algorithm for university students' psychological fitness and performance detection," *Heliyon,* vol. 9, no. 8, p. e18550, 2023. https://doi.org/10.1016/j.heliyon.2023.e18550

[35] A. W. Akram and Z. Alamgir, "Distributed fuzzy clustering algorithm for mixed-mode data in Apache SPARK," *J. Big Data,* vol. 9, no. 1, 2022. https://doi.org/10.1186/s40537-022-00671-7

[36] S. Patil and H. Solanki, "Feasibility of fuzzy clustering for improving the objective function of K-means clustering," *International Journal of Engineering Research and Technology (IJERT)*, vol. 11, no. 7, pp. 489–493, 2022. https://doi.org/10.17577/IJERTV11IS070231

[37] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler, "Exploring topic coherence over many models and many topics," in *Conference on Empirical Methods in Natural Language Processing*, 2012, pp. 952–961.

[38] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Edinburgh, Scotland, UK, 2011, pp. 262–272.

[39] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics,* vol. 20, pp. 53–65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7

[40] S. Myagmarsuren, "Exploring the use of silhouette score in K-means clustering for image segmentation," *International Journal of Engineering Research and Technology (IJERT),* vol. 13, no. 4, 2024. https://doi.org/10.17577/IJERTV13IS040316

[41] X. Wu, L. A. T. Nguyen, and A.T. Luu, "A survey on neural topic models: Methods, applications, and challenges," *Artificial Intelligence Review,* 2024. https://doi.org/10.21203/rs.3.rs-3049182/v1

[42] F. Nan, Ran Ding, Ramesh Nallapati, and Bing Xiang, "Topic modeling with Wasserstein autoencoders," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, 2019, pp. 6345–6381. https://doi.org/10.18653/v1/P19-1640

[43] A. B. Dieng, F. J. R. Ruiz, and D. M. Blei, "Topic modeling in embedding spaces," *Ransactions of the Association for Computational Linguistics,* vol. 8, pp. 439–453, 2019. https://doi.org/10.1162/tacl_a_00325

[44] Z.-Y. Lim, L.-Y. Ong, and M.-C. Leow, "A review on clustering techniques: Creating better user experience for online roadshow," *Future Internet,* vol. 13, no. 9, p. 233, 2021. https://doi.org/10.3390/fi13090233

[45] B. A. Jose and P. Das, "A multi-objective approach for inter-cluster and intra-cluster distance analysis for numeric data," in *Soft Computing: Theories and Applications*, in Lecture Notes in Networks and Systems, R. Kumar, C. W. Ahn, T. K. Sharma, O. P. Verma, and A. Agarwal, Eds., Singapore, Springer Nature Singapore, 2022, pp. 319–332. https://doi.org/10.1007/978-981-19-0707-4_30

[46] J. Soler, F. Tencé, L. Gaubert, and C. Buche, "Data clustering and similarity," in *International Florida Artificial Intelligence Research Society Conference*, 2013, pp. 492–495. https://cdn.aaai.org/ocs/5843/5843-29829-1-PB.pdf

[47] D. D. Burns, "Feeling good: The new mood therapy, william morrow and company," Wikipedia, 2024. https://en.wikipedia.org/wiki/Feeling_Good:_The_New_Mood_Therapy

[48] J. S. Beck, *Cognitive Behavior Therapy: Basics and Beyond* (2nd ed.). New York, NY: Guilford Press, 2011. https://psycnet.apa.org/record/2011-22098-000

[49] P. McCarthy and S. Jarvis, "vocd: A theoretical and empirical evaluation," *Language Testing*, vol. 24, no. 4, pp. 459–488, 2007. https://doi.org/10.1177/0265532207080767

[50] S. Mifrah and E. H. Benlahmar, "Topic modeling with transformers for sentence-level using coronavirus corpus," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 17, pp. 50–59, 2022. https://doi.org/10.3991/ijim.v16i17.33281

[51] A. Al-zubidi, N. F. AL-Bakri, R. K. Hasoun, S. Hassan Hashim, and H. T. Alrikabi, "Mobile application to detect Covid-19 pandemic by using classification techniques: Proposed system," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 15, no. 16, pp. 34–51, 2021. https://doi.org/10.3991/ijim.v15i16.24195

[52] R. Alnusyan, R. Almotairi, S. Almufadhi, A. A. Al-Shargabi, and J. F. Alshobaili, "Hybrid approach for user reviews' text analysis and visualization: A case study of amazon user reviews," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 8, pp. 79–93, 2022. https://doi.org/10.3991/ijim.v16i08.30169

8 AUTHORS

Marwah Alian is a faculty member at the Computer Science Department, Faculty of Information Technology, The World Islamic Sciences and Education University, Amman, Jordan. Also, she was a member of the Basic Sciences Department, Faculty of Science, The Hashemite University, Zarqa, Jordan, till Feb 2024. She received her B.Sc. in Computer Science from The Hashemite University in 1999. Then, she received M.Sc. degree in Computer Science from The University of Jordan in 2007. In 2021, she received her PhD from Princess Sumaya University for Technology. She has a number of publications in e-learning systems, data mining, mobile applications, adaptive learning, mobile learning, and natural language processing (E-mail: Marwah.alian@wise.edu.jo, Marwahm@hu.edu.jo).

Maisa Al-Khazaleh is a faculty member at the Basic Sciences Department, Faculty of Science, Hashemite University, Zarqa, Jordan. She received her B.Sc. degree in Computer Science from Jordan University of Science and Technology in 2007, and then she received the M.Sc. degree in Computer Science from Yarmouk University in 2009. Her research interests are primarily in the area of machine learning and natural language processing (E-mail: Maisa@hu.edu.jo).

Hadeel AlShboul is a full-time Lecturer at Computer Science Department, Faculty of Information Technology, The World Islamic Sciences and Education University (Jordan). She received her B.Sc. degree in Computer Science from AlBalqa Applied University in 2020, and then she received her M.Sc. degree in Computer Science from Princess Sumaya University for Technology in 2022. Her research interests include Machine Learning, Artificial Intelligence, and E-learning systems (E-mail: hadeel.alshboul@wise.edu.jo).