Coping with ADHD in College: Topic Modeling and Sentiment Analysis of Online Peer Discussions on Reddit

Sihan Jian
Florida State University
sjian@fsu.edu

Saptarshi Bhowmik
Florida State University
sbhowmik@fsu.edu

Bret Staudt Willet
Florida State University
bret.staudtwillet@fsu.edu

ABSTRACT

Navigating college with Attention-Deficit/Hyperactivity Disorder (ADHD) presents unique academic and personal challenges, often leaving students seeking peer support beyond formal institutional resources. This study applies latent Dirichlet allocation (LDA) topic modeling, sentiment analysis, and emotion classification to analyze 793 posts and 4,911 comments in r/adhd_college, the largest academic ADHD community on Reddit. Findings reveal six dominant topics: academic performance, physical and emotional well-being, study management strategies, ADHD medication and effects, accommodations and support services, and research participation. While discussions generally exhibit a positive and supportive tone, sentiment variability highlights frustration surrounding academic stress, mental health, medication management, and institutional support. The prominence of positive sentiment and gratitude in comment interactions emphasizes the role of r/adhd_college as a safe, peer-driven support space offering members practical advice and coping strategies. By unveiling the unfiltered perspectives of ADHD college students, this study contributes to a deeper understanding of their lived experiences and provides insights for designing community-based resources for academic success and overall well-being.

Keywords

College Students, ADHD, Topic Modeling, Sentiment Analysis, Emotion Classification, Social Media, Reddit

1. INTRODUCTION

Attention-Deficit/Hyperactivity Disorder (ADHD) is a neurodevelopmental condition characterized by inattention, hyperactivity, and impulsivity, often persisting into adulthood. Among college students, ADHD is notably prevalent, with 2% to 8% self-reporting significant symptoms [20] and about 25% of students using disability services diagnosed with ADHD [5]. These students frequently struggle with organization, time management, and planning (OTP), which can

negatively affect their academic performance and social interactions [9]. To manage these challenges, many seek support in online communities, where they connect with peers, share strategies, and develop a better understanding of their experiences [4].

Social media, rich with user-generated content, offers a valuable resource for studying ADHD in naturalistic settings. Reddit, a platform built around themed subreddits and anonymous discussions, provides a space for open conversations on sensitive topics like ADHD. Unlike formal surveys or clinical studies, these informal exchanges offer a unique opportunity to uncover authentic experiences and practical strategies shared by community members. Prior data mining studies on ADHD-related online communities have mainly focused on linguistic patterns and diagnostic models [6, 2, 11], often comparing ADHD users with other neurodivergent groups [8] or examining subgroups within the ADHD community [13]. However, less attention has been given to how ADHD college students use online spaces to navigate academic and social challenges. The subreddit r/adhd_college (https://www.reddit.com/r/adhd_college/) exemplifies this phenomenon, providing an affinity space for college students and continuing education learners with ADHD to connect and navigate their shared experiences and academic and personal challenges associated with ADHD. Within this digital community, members engage in peer-to-peer discussions that foster knowledge exchange and mutual support, free from the formalities of traditional educational or clinical settings. As the largest academic ADHD community on Reddit, r/adhd_college provides a valuable window for understanding the lived experiences of individuals with ADHD in higher education.

In natural language processing (NLP), topic modeling and sentiment analysis are widely used to analyze large volumes of social media text. Topic modeling, a machine learning technique, identifies and characterizes hidden themes within vast text corpora using Bayesian inference to uncover underlying structures [3]. Latent Dirichlet Allocation (LDA) is one of the most commonly used algorithms in social media research, with studies demonstrating its effectiveness in extracting meaningful topics from user-generated content [1]. In education research, LDA has been applied to explore online discourse, such as analysis of TikTok content on engineering education [17] and dynamic topic modeling of an AI painting subreddit, showcasing its versatility in informal learning contexts [19]. These studies highlight LDA's po-

tential for examining students' perspectives in digital spaces. Sentiment analysis, a key area of affective computing, extracts sentiments from various human expressions, including text, speech, and movement [21]. Social media platforms serve as rich sources of sentiment-laden data, with studies demonstrating the value of combining LDA with sentiment analysis to identify mental health concerns [14] and track patterns in university students' online discourse [12]. Traditional sentiment analysis often focuses on opinion mining, classifying text into positive, negative, or neutral categories. However, sentiment encompasses more than just opinion, as emotions can shape or be shaped by expressed viewpoints [21]. Everyday emotions, such as gratitude, love, frustration, and confusion, do not always fit neatly into these broad categories and often convey more nuanced feelings. This distinction is crucial in ADHD-related discussions, where capturing subtle emotions can offer deeper insights into the challenges, frustrations, and coping mechanisms of college students with ADHD.

By integrating sentiment analysis with emotion classification, this study aims to capture the affective dimensions of discussions within the ADHD college student community, offering a more holistic view of their academic and personal experiences. Rather than limiting analysis to individual posts and comments, we also examine sentiment and emotion at the topic level to reveal broader emotional patterns that shape community interaction. This multi-layered approach helps illuminate how members articulate and respond to various concerns and experiences, providing a richer understanding of the community's emotional landscape. Building on this approach and addressing gaps in the literature, the study focuses on the r/adhd_college subreddit and explores the following research questions:

RQ1: What are the most common topics discussed within the ADHD college student community?

RQ2: What is the overall sentiment in the data, and how does it vary across posts, comments, and topics?

RQ3: What emotions are most prevalent in the data, and how do they vary across posts, comments, and topics?

2. METHODS

2.1 Data Collection
This study focuses on r/adhd_college, the largest academic ADHD community on Reddit, which had approximately eleven thousand members at the time of data collection in December 2024. Data were collected using Asynchronous Python Reddit API Wrapper (Async PRAW) to interface with Reddit data, allowing us to retrieve historical posts and comments from the community. The dataset includes all posts and comments made between the community's inception on November 6, 2020, and December 29, 2024, totaling 793 posts and 4,911 comments. Prior to data collection, we submitted a Determination Form for Exempt or Human Subjects Research and received an exemption from the Institutional Review Board (IRB), as the study was classified as non-human subjects research.

2.2 Data Preprocessing

We began by concatenating each post's title, text, and comments into a single document, creating a unified representation of each thread for a more comprehensive analysis. To prepare the data, we implemented a multi-step preprocessing pipeline using the Natural Language Toolkit (NLTK). First, we tokenized the text, removing punctuation and special characters, expanded contractions for clarity, and replaced emojis with spaces for standardization. We then removed both generic and dataset-specific stop words before applying lemmatization with the WordNet lemmatizer to normalize word variations. Finally, we discarded posts with empty or invalid text after preprocessing to maintain data integrity. This structured approach optimized the dataset for NLP techniques, ensuring robust topic modeling and sentiment analysis.

2.3 Data Analysis and Visualization
2.3.1 Topic Modeling
We applied LDA to identify underlying themes within the dataset, implementing the model using the Gensim library. To determine the optimal number of topics, we conducted an initial exploration ranging from 1 to 20, followed by a refined search from 20 to 30. We adjusted key parameters to enhance model stability and optimize topic distribution across documents. To assess model performance, we used coherence scores, which measure the semantic similarity among high-probability words within a topic, indicating how well the topics represent coherent themes [18]. Specifically, we employed the c_v coherence metric and visualized coherence scores across different topic configurations. We observed local extrema between 6 and 7 topics, with coherence scores of 0.382 and 0.381, respectively. Comparing the interpretability of both models, we selected six topics as the optimal number. Once the LDA model was trained, we assigned a dominant topic to each thread entry by computing the topic distribution for each document and selecting the topic with the highest probability. We then incorporated these topic assignments into the dataset as an additional column, providing a structured basis for further sentiment analysis.

To enhance topic interpretation, we used the LDAvis package [16] to generate an interactive visualization. This tool displayed the 30 most salient terms for each topic, helping to identify key themes and explore relationships between topics. We manually reviewed the top words within each topic to define overarching themes, assigning descriptive labels based on the most representative terms. This final step was crucial in translating the LDA model's numerical outputs into meaningful insights that accurately reflect the thematic composition of the dataset.

2.3.2 Sentiment Analysis and Emotion Classification
To capture both overall sentiment polarity and nuanced emotional expressions in the textual data, we implemented sentiment analysis coupled with emotion classification to explore how different topics were perceived and discussed. For sentiment analysis, we used the VADER module in NLTK [7], a lexicon-based tool optimized for social media text that calculates sentiment polarity using compound scores. We extracted VADER compound scores for each text entry and analyzed sentiment distribution by computing the median and interquartile range (IQR) at the topic level. These dis-

tributions were visualized using box plots to show sentiment variability across topics.

To examine emotional expressions more deeply, we applied the Roberta-base-go_emotions model from Hugging Face, a RoBERTa-based deep learning model fine-tuned on Reddit data to classify 28 discrete emotion categories [10]. This model generated predicted emotions for each text entry, along with corresponding confidence scores. To enhance interpretability, we processed the extracted emotions through several key steps. First, we standardized the data structure by cleaning and flattening the emotion output. Next, we aggregated emotions at the topic level by computing the average emotion scores across all entries within each topic. Finally, we visualized the emotional intensity within the data using bar charts and heatmaps, highlighting the distribution of emotions in posts, comments, and across all data per topic.

3. RESULTS

3.1 RQ1: LDA Topic Modeling

The topic modeling results identified six key themes commonly discussed within the community. As shown in Table 1, these distinct topics emerged after adjusting the relevance metric (λ). Topic 1, the most prominent, centers on academic performance and navigating college challenges, with keywords such as "class", "professor", "failed". Topic 2 focuses on physical and emotional well-being, highlighted by terms such as "workout", "insomnia", and "crisis". Topic 3 pertains to study management tools and strategies, featuring words such as "note", "planner", and "app". Topic 4 revolves around ADHD medications and their effects on daily life and diet, as indicated by keywords such as "medication", "dose", and "adderall". Topic 5 encompasses discussions on accommodations and disability support services, with terms such as "accommodation", "disability", and "extension". As illustrated in Figure 1, Topic 1 is closely related to Topic 5, while Topics 2 and 4 notably overlap and both connect to Topic 1. Topic 3 has a minor overlap with Topic 4. Topic 6 stands apart, focusing on research participation and recruitment, with keywords such as "survey", "participant", and "consent".
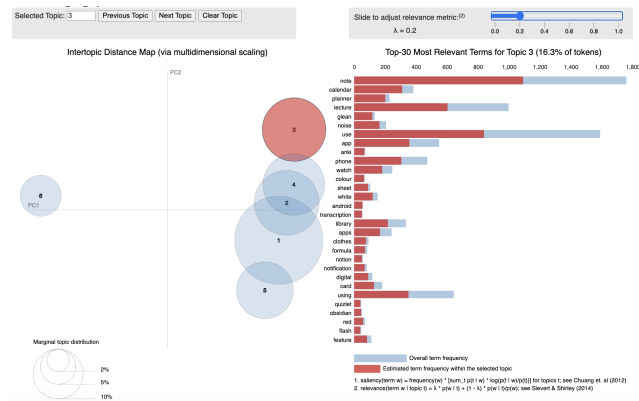

Figure 1: Intertopic distance map with the top-30 most relevant terms

3.2 RQ2: Sentiment Polarity

Using user post data, the distribution of sentiment scores reveals a clear trend of positive sentiment, as indicated by the higher concentration of positive sentiment scores (see Figure 2). Out of 793 posts, 466 (58.8%) exhibited positive sentiment, 225 (28.4%) reflected negative sentiment, and 102 (12.9%) were neutral. The median sentiment score is 0.61, with an IQR of 1.31. This suggests that while community posts generally maintain a positive tone, they are also characterized by a broad range of emotional expressions, including occasional strong negative sentiments and limited neutral discourse. The wide variability in sentiment scores reflects the diverse emotional experiences expressed in the user posts within the community.

In comparison, when analyzing the scatterplot of user comments, the sentiment scores are more tightly concentrated on the positive side (see Figure 3). Among the 4,911 comments, a substantial majority of 3,369 comments (68.6%) expressed positive sentiment, while 878 comments (17.9%) conveyed negative sentiment, and 664 comments (13.5%) were neutral. While positive sentiments still dominate, there is a noticeable reduction in the variability of sentiment scores compared to posts. The median sentiment score for comments is 0.49, with a narrower IQR of 0.83, indicating less emotional fluctuation within comments.
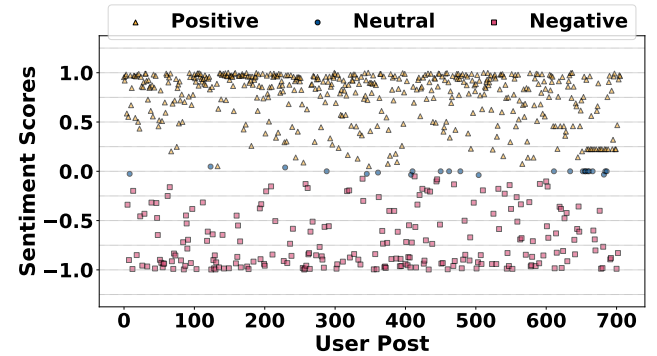

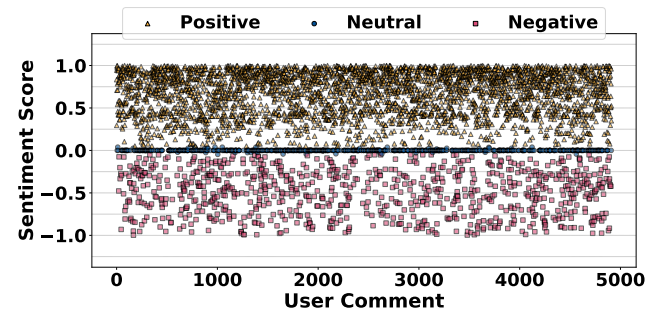**Figure 2: Sentiment distribution of user posts**

Figure 3: Sentiment distribution of user comments

Consistent with the analyses of posts and comments separately, the distribution of sentiment scores across the six identified topics using aggregated thread data reveals a clear

Table 1: Identified topics and keywords

| Topic ID	Topic	Relevance Metric (λ)	Keyword	Token Contribution (%)
1	Academic Performance and Challenges	0.28	class, professor, failed, college, phd, credit, program, learning, degree, scholarship	31.5%
2	Physical and Emotional Well-being	0.08	workout, insomnia, crisis, nighter, melatonin, motivating, wake, pray, pissed, intimidating	17%
3	Study Management Tools and Strategies	0.20	note, planner, app, calendar, glean, anki, phone, transcription, notification, quizlet	16.3%
4	ADHD Medication and Effects	0.35	medication, dose, adderall, stimulant, food, eat, price, stressful, strattera, brain	15.4%
5	Accommodations and Disability Support Services	0.40	accommodation, disability, extension, ask, exam, extra, testing, diagnosis, ada, documentation	13%
6	Research Participation and Recruitment	0.20	survey, participant, consent, qualtrics, consent, research, questionnaire, ethical, eligibility, anonymous	6.4%
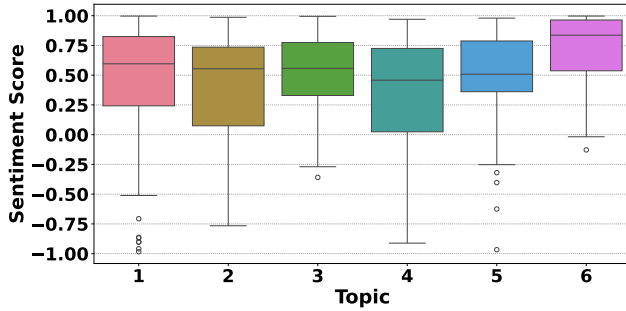
Figure 4: Sentiment distribution by topic

trend of positive sentiment within the community (see Figure 4), with all topics showing median sentiment scores above zero. Research participation and recruitment (Topic 6) stands out with the highest median sentiment score and the least variability, indicating discussions that are not only overwhelmingly positive but also consistently so. In contrast, topics such as academic performance and challenges (Topic 1), physical and emotional well-being (Topic 2), and ADHD medication and effects (Topic 4) display greater variability in sentiment scores, with noticeable negative outliers in academic performance and challenges (Topic 1), reflecting a broader range of emotional experiences. Meanwhile, study management tools and strategies (Topic 3) and accommodations and disability support services (Topic 5) maintain consistently positive sentiment with moderate variability. While occasional negative sentiments emerge, particularly regarding accommodations, the overall tone remains positive.

3.3 RQ3: Emotion Classification
As shown in Figure 5, gratitude is especially prominent in comments, indicating frequent expressions of appreciation, while posts are primarily neutral with slightly higher levels of gratitude and love. Using aggregated thread data, the distribution of emotion scores in Figure 6 reveals that

positive emotions, particularly gratitude, amusement, admiration, and optimism, are the most prevalent within the community. Emotion scores vary across topics, with neutral emotion being dominant in research participation and recruitment (Topic 6). In contrast, physical and emotional well-being (Topic 2) shows slightly higher levels of disgust, while academic performance and challenges (Topic 1) and ADHD medication and effects (Topic 4) display slightly elevated levels of fear. Overall, while negative emotions such as remorse, disgust, and fear are present, they occur less frequently than positive emotions.
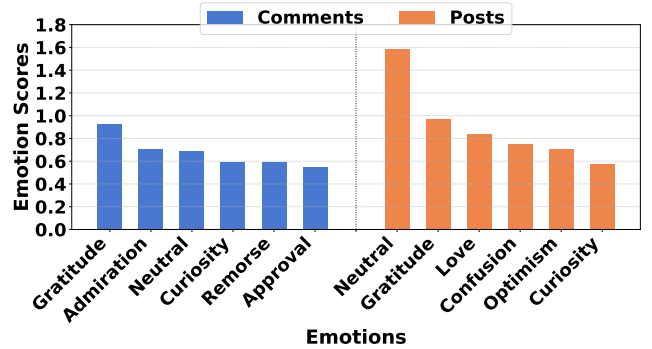

Figure 5: Emotion comparison between comments and posts

4. DISCUSSIONS
4.1 Challenges of College Students with ADHD
The topics discussed within r/adhd_college align with existing research on ADHD in college students [20, 5]. Compared to their neurotypical peers, students with ADHD often have greater concerns about their academic abilities [15]. Our topic modeling results reflect these challenges, with academic struggles emerging as the dominant theme, highlighting the centrality of academic concerns for r/adhd_college members, many of whom likely self-identify as having ADHD. Despite the generally positive tone revealed by sentiment
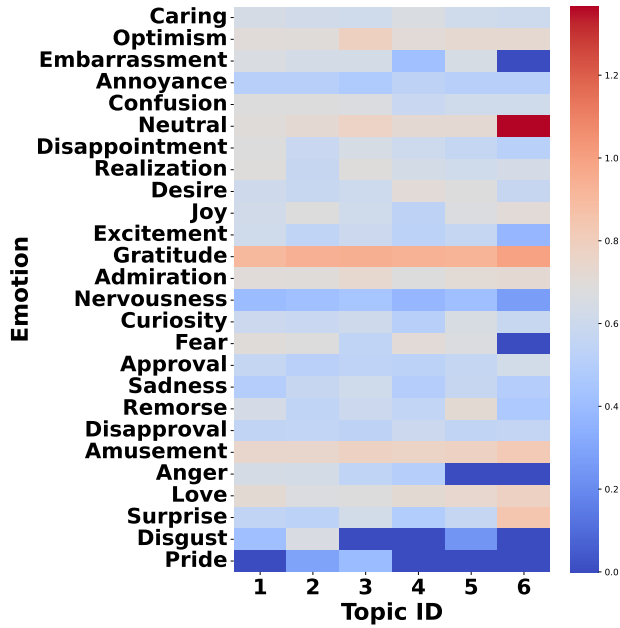
Figure 6: Emotion classification by topic

analysis, discussions on academic stress, mental health, medication, and institutional support show wide sentiment variation, with occasional strongly negative tones. Elevated levels of fear and disgust in these discussions suggest frustration, uncertainty, and criticism, which further indicates that these are likely the key challenges faced by ADHD college students. To better support ADHD students in their transition to and navigation of college life, higher education institutions should consider these challenges when developing targeted interventions and support services.

4.2 Online Communities as Sources of Support for ADHD Students

The predominantly positive sentiment in r/adhd_college reflects the community's supportive nature. While posts express a range of emotions, comments are consistently more positive, often conveying gratitude, which suggests an affirming and encouraging dynamic. Posts typically share personal experiences, while responses offer advice and validation, fostering a sense of support. A distinct focus on study tools and strategies further highlights the community's solution-oriented discussions. These findings demonstrate that r/adhd_college serves as both a source of emotional support and a hub for practical strategies, helping ADHD students navigate academic challenges. This reinforces the insights of [4] that online communities provide a safe space for young adults with ADHD to share experiences and connect with others, with anonymity further strengthening this support. Higher education institutions can explore ways to integrate online communities into their support frameworks for ADHD students.

4.3 Limitations and Future Research

One major limitation of this study is the relatively small dataset and the lack of demographic information, due to the anonymity and data privacy policies of platforms like Reddit. These constraints limit the generalizability of our findings. Future research could address this limitation by expanding the dataset to include additional social media platforms and validating the findings through empirical studies that explore the experiences of ADHD college students in both online and offline contexts. Although previous research has identified malingering and medication misuse as concerns among college students with ADHD [5], these topics were largely absent from r/adhd_college discussions. This gap may reflect limitations in our topic modeling approach, which may struggle to detect more nuanced or sensitive subtopics. Future work could refine these techniques for greater granularity, incorporate supervised learning to improve classification accuracy, and explore advanced models that account for topic interdependencies to better capture semantic overlaps.

5. CONCLUSION

This study used LDA topic modeling to examine discussions from a prominent online academic community for college students with ADHD, identifying core themes such as academic performance, emotional and physical well-being, study strategies, medication effects, accommodation services, and research participation. The sentiment and emotion patterns observed across these topics not only highlight the challenges faced by ADHD students but also demonstrate the community's role in providing support and practical advice. Additionally, our findings point to the value of combining opinion mining with emotion classification to uncover richer emotional dynamics in online discourse, presenting a promising avenue for future research.

6. REFERENCES

[1] R. Albalawi, T. H. Yeap, and M. Benyoucef. Using topic modeling methods for short-text data: A comparative analysis. *Frontiers in artificial intelligence*, 3:42, 2020.

[2] N. Alsharif, M. H. Al-Adhaileh, S. N. Alsubari, and M. Al-Yaari. Adhd diagnosis using text features and predictive machine learning and deep learning algorithms. *Journal of Disability Research*, 3(7):20240082, 2024.

[3] R. Churchill and L. Singh. The evolution of topic modeling. *ACM Computing Surveys*, 54(10s):1–35, 2022.

[4] C. M. Ginapp, N. R. Greenberg, G. Macdonald-Gagnon, G. A. Angarita, K. W. Bold, and M. N. Potenza. The experiences of adults with adhd in interpersonal relationships and online communities: A qualitative study. *SSM-Qualitative Research in Health*, 3:100223, 2023.

[5] A. L. Green and D. L. Rabiner. What do we really know about adhd in college students? *Neurotherapeutics*, 9(3):559–568, 2012.

[6] S. C. Guntuku, J. R. Ramsay, R. M. Merchant, and L. H. Ungar. Language of adhd in adults on social media. *Journal of attention disorders*, 23(12):1475–1485, 2019.

[7] C. Hutto and E. Gilbert. Vader: A parsimonious

rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225, 2014.

[8] N. Kalantari, A. Payandeh, M. Zampieri, and V. G. Motti. Understanding the language of adhd and autism communities on social media. In *2023 IEEE International Conference on Big Data (BigData)*, pages 2188–2195. IEEE, 2023.

[9] J. M. Langberg, J. N. Epstein, and A. J. Graham. Organizational-skills interventions in the treatment of adhd. *Expert review of neurotherapeutics*, 8(10):1549–1561, 2008.

[10] S. Lowe. roberta-base-go_emotions (revision 58b6c5b), 2024.

[11] S. M. Mumu, H. Hoque, and N. Sakib. A classified mental health disorder (adhd) dataset based on ensemble machine learning from social media platforms. In *Proceedings of the Fourth International Conference on Trends in Computational and Cognitive Engineering: TCCE 2022*, pages 395–405. Springer, 2023.

[12] F. Nwaoha, Z. Gaffar, H. J. Chun, and M. Sokolova. Longitudinal sentiment topic modelling of reddit posts. *arXiv preprint arXiv:2401.13805*, 2024.

[13] M. M. Rahman. Linguistic dynamics: Women vs. the general population in reddit's adhd discussions. *medRxiv*, pages 2024–09, 2024.

[14] K. Rosamma and K. Rosamma Jr. Analyzing online conversations on reddit: A study of stress and anxiety through topic modeling and sentiment analysis. *Cureus*, 16(9), 2024.

[15] B. Shaw-Zirt, L. Popali-Lehane, W. Chaplin, and A. Bergman. Adjustment, social skills, and self-esteem in college students with symptoms of adhd. *Journal of attention disorders*, 8(3):109–120, 2005.

[16] C. Sievert and K. Shirley. Ldavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, pages 63–70, 2014.

[17] S. Solanki, M. A. Tsugawa, H. Karimi, et al. Leveraging social media analytics in engineering education research. In *2023 ASEE Annual Conference & Exposition*, 2023.

[18] S. Syed and M. Spruit. Full-text or abstract? examining topic coherence scores using latent dirichlet allocation. In *2017 IEEE International conference on data science and advanced analytics (DSAA)*, pages 165–174. Ieee, 2017.

[19] S. Wei and R. Bi. Uncovering the evolution of topics about ai painting: Dynamic topic modeling of 180k discourse data in an online community. In *Proceedings of the 17th International Conference on Educational Data Mining*, pages 672–678, 2024.

[20] L. L. Weyandt and G. DuPaul. Adhd in college students. *Journal of attention disorders*, 10(1):9–19, 2006.

[21] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane. Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):1–33, 2017.