Analyzing Primary Sources of Negative Stress in Adolescents Using LSTM Models from Social Media Data

Pragathi Ashok, Dr. Neha Keshan

Rensselaer Polytechnic Institute, 110 8th St, Troy, NY 12180

Abstract. The levels of stress experienced by adolescents worldwide represent a pressing and increasingly significant issue. Prior studies have employed various sentiment analysis techniques to identify mental health disorders such as depression and anxiety related to a wide variety of individuals. In this study, we leverage Long Short-Term Memory (LSTM) models to analyze and classify stressors specific to adolescents aged 10–19. Using Kaggle datasets sourced from social media platforms like Reddit and Twitter, we examine four primary causes of negative stress. Not much research has utilized sentiment analysis specifically for stress detection in adolescents, making this study unique in its focus. By addressing this gap, our research contributes to a deeper understanding of adolescent stress through the analysis of social media data. By identifying these stressors, we aim to provide insights that can guide the development of targeted intervention strategies, such as mental health campaigns and support services. These efforts could significantly reduce stress levels and improve the overall mental well-being of adolescents. A link to the Github Repo with all relevant files is included: [6].

1 Introduction

Adolescents worldwide face alarmingly high stress levels, affecting their mental and physical health [5]. They struggle to balance academic demands, societal pressures, changing friendships, and uncertainties about the future [15]. Many turn to social media for connection or a temporary escape [16]. Platforms like TikTok, Twitter, Instagram, and Reddit offer spaces for self-expression, shared experiences, and community building. The user-driven content on these platforms reflects emotions and behaviors at scale, making it a valuable resource for analyzing negative adolescent stress through sentiment analysis. Negative stress, in this context, refers to feelings of being overwhelmed, anxious, or worried due to challenging situations [11].

This stress can be uncovered using sentiment analysis, a branch of Natural Language Processing (NLP), that extracts opinions, emotions, and sentiments from text and classifies them as positive, neutral, or negative [14]. Traditional machine learning models, such as support vector machines and decision trees, have been widely used for this task but often struggle with the complexities of human language. Advanced deep learning techniques, like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformers, are more effective at capturing context and relationships within data [14].

This study uses LSTM networks, an advanced type of RNN. LSTMs excel at processing sequential data by retaining relevant information and discarding unnecessary details, enabling more accurate predictions [7]. By analyzing text from platforms like Reddit and Twitter, this study employs LSTMs to identify and categorize sources of negative adolescent stress.

Hence, in this study, we aim to use sentiment analysis to uncover and classify negative stress expressed by adolescents on social media. The goal is to provide insights that can inform strategies to address and mitigate this growing issue.

2 Motivation

I was motivated to conduct this study because I have experienced the effects of negative stress while navigating the challenges of growing up in a fast-paced, ever-changing world. While some stress can be motivating, such as striving to excel in academics or preparing for a major career opportunity, excessive levels often diminish quality of life. They can make routine activities, like attending school, maintaining relationships, or even engaging in hobbies, feel overwhelming.

In addition to this personal experience, I became particularly intrigued by deep learning techniques after reviewing a paper titled, "Aspect based sentiment analysis using deep learning approaches: a survey." This paper provided a comprehensive review of deep learning approaches for aspect-based sentiment analysis (ABSA), highlighting their effectiveness in capturing sentiment related to specific aspects within text. What I found particularly intriguing about deep learning techniques from this paper, was their ability to combine sequential text processing with real-world knowledge for more accurate sentiment analysis. For example, when LSTMs are paired with SenticNet, they not only capture the flow of text but also leverage concepts like associating "pizza" with "food quality," significantly improving the accuracy of aspect extraction in complex reviews [10]. I found this information to be deeply interesting, and wanted to explore more about these techniques.

Hence, I undertook this project as a way of exploring my interests in deep learning while applying my skills to address a real-world issue—the growing impact of negative stress and its consequences on adolescents.

3 Literature Review

To build a strong foundation for this research, a review of relevant literature was conducted, including papers from reputable sources like Google Scholar.

3.1 Sentiment Analysis Technologies

One paper, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review," highlighted the advantages of deep learning models like RNNs and CNNs over traditional methods such as SVM and Naïve Bayes, which struggle with large datasets and nuanced emotions [8]. These findings supported the selection of LSTMs as the primary sentiment analysis technique for this study. Another paper, "Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions," further emphasized the effectiveness of deep learning methods, by highlighting their ability to process large-scale data and capture subtle emotional patterns that traditional models often miss. The paper explored the strengths of RNNs, CNNs, and hybrid approaches in identifying complex relationships within textual data.

3.2 Addressing Challenges Related to LSTM Technologies

In addition to collecting information on different types of technologies, information was also gathered to address some of the challenges associated with these technologies. The paper "A Comprehensive Review of Visual–Textual Sentiment Analysis from Social Media Networks" discussed multimodal sentiment analysis, demonstrating how combining textual and visual data enhances model accuracy. It highlighted challenges like managing noisy data and adapting models to new domains, often requiring transfer learning or domain-specific fine-tuning to maintain performance across platforms [4]. Similarly, "A review of sentiment analysis: tasks, applications, and deep learning techniques" addressed the importance of tailoring models to different domains to achieve consistent results [14]. These insights prepared us to anticipate obstacles during implementation while reinforcing the decision to focus on LSTMs for textual data analysis.

3.3 Real World Applications of Sentiment Analysis

In addition to technical considerations, literature also emphasized the importance of interpreting sentiment analysis results. The paper "Stress Detection from Social Media Articles: New Dataset Benchmark and Analytical Study" shifted attention towards interpreting the insights generated by sentiment analysis rather than comparing model performances. This perspective informed the study's aim to focus on understanding stress sources and patterns within adolescent social media posts [13].

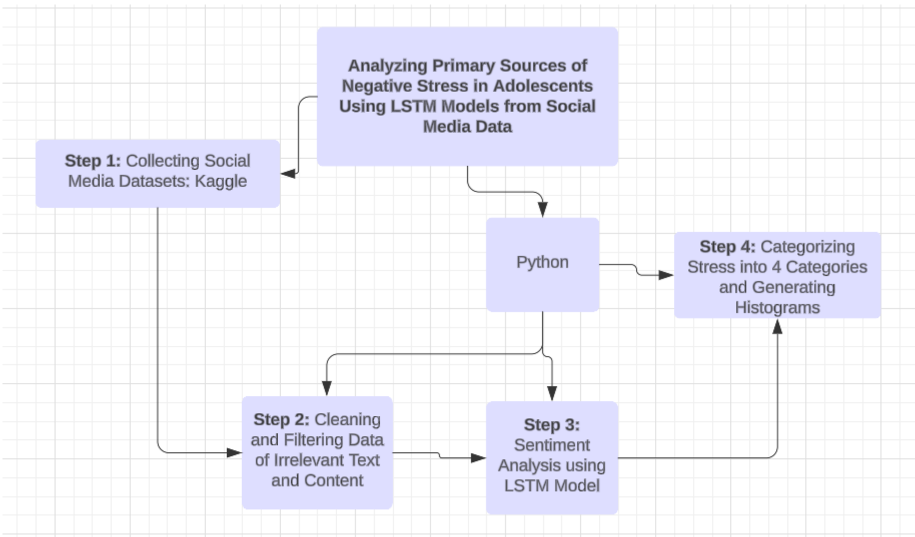
Lastly, insights on applying NLP to real-world issues guided this study's practical approach. The paper "Using Natural Language Processing to Explore Social Media Opinions on Food Security: Sentiment Analysis and Topic Modeling Study" demonstrated how NLP could inform evidence-based interventions when combined with expert interpretations. This inspired the strategy of consulting psychologists and relevant experts to develop effective mitigation strategies. [12].

Hence, through this comprehensive literature review, LSTM models emerged as the most suitable sentiment analysis technique for this study. Furthermore, the review revealed a gap in research on using sentiment analysis to explore adolescent stress expressed on social media, highlighting an opportunity to make meaningful contributions in this area.

4 Research Problem/Contribution

This study focuses on using LSTM models to analyze the primary sources of negative stress expressed by adolescents (ages 10–19) on social media platforms like Reddit and Twitter. The objectives are to detect adolescent stress in existing social media datasets, identify whether it is negative, classify it into four main categories, compare stress levels in each of these categories between Reddit and Twitter, and propose strategies to help adolescents better manage their stress.

5 Methods and Data



The completion of this study involved several key steps: gathering data, cleaning and filtering it, visualizing the results, and conducting quantitative analysis.

5.1 Collecting Data Using Kaggle

To collect the data, the free website Kaggle, which offers a wide variety of datasets for research purposes, was utilized. Relevant datasets were identified by entering keywords such as "social media," "sentiment analysis," and "stress." Using these search terms, five relevant datasets were obtained, focusing on mental health classification on Twitter and Reddit (see Table 9). This data served as the foundation for performing the sentiment analysis.

5.2 Cleaning and Filtering Data of Irrelevant Text and Content

The next challenge was to clean the datasets—transforming raw, messy data into a meaningful structure. A Python script was crafted to remove irrelevant text like URLs, special characters, and extra spaces, ensuring only relevant content remained. Filtering went a step further to target posts specifically related to adolescent stress. Four essential categories and terms relevant to each category were defined. These categories were stress terms, educational terms, age terms, and slang terms (see table 9). Using pattern-matching techniques, each text entry was evaluated based on how many keywords it matched in each category. Based on how many terms were identified in each category, a score was generated (between 0-1) for each text entry.

5.3 Relevance Score Calculation

The calculation is performed as follows: in each text entry, the number of matching terms identified in each of the 4 categories are counted. These numbers are divided by 2 to generate a base score for each category (with scores capped at 1.0, meaning you need at least 2 terms to achieve the maximum score in any category). These category scores are then weighted according to their importance, with stress and age terms each contributing 30 percent to the final score (reflecting their primary importance in identifying adolescent stress), while education and slang terms each contribute 20 percent (serving as supporting contextual indicators).

The weighted scores are summed to produce a final relevance score between 0 and 1, rounded to three decimal places. Any text entry scoring 0.5 or higher is retained in the dataset, ensuring that only the most relevant content is kept for analysis. The cleaned datasets retain the original columns and include additional ones, such as the keywords identified for each category, their respective counts, and the calculated relevance scores.

5.4 Classifying Sentiments Using LSTM Model

Once the datasets were cleaned, the next step was to classify the sentiments of each text analyzed into three categories using an LSTM model: Positive, Neutral, and Negative (labeled 0, 1, and 2 respectively). In order to do this, the cleaned Twitter datasets were combined into one dataset and the cleaned Reddit datasets into another. The data was then split into training and testing sets, and Textblob, a python library, was applied to the training set to assign initial sentiment labels to each text entry based on polarity scores. The LSTM then uses this data to learn how to classify different text entries, and is applied to the testing data to see how well it generalizes.

The negative sentiment predictions for Reddit and Twitter are exported to 3 separate datasets, one for each platform (Reddit and Twiiter), and one with the combined results. The datasets include insights such as the sentiment prediction (2 for negative), the confidence score, and the probabilities of each sentiment classification (postive, negative, and neutral) of each text entry.

5.5 LSTM Model Layers:

The LSTM model, the heart of this analysis, comprises three main layers: an embedding layer, LSTM layers, and a dense layer. The embedding layer transforms each word into a dense numerical vector, capturing its meaning. The LSTM layers identify and retain word patterns across sequences to interpret the flow of text, which is essential for sentiment analysis. Finally, the dense layer, with softmax activation, outputs probability scores for each sentiment class—positive, neutral, or negative (see Table 9). Each text entry is assigned the sentiment class with the highest probability score, which also serves as the confidence score. For instance, if the positive sentiment has the highest probability, the entry is labeled as positive. The model refines its predictions by minimizing errors over several training epochs.

5.6 Categorizing and Visualizing Data Using Histograms

The final step involved categorizing the data using visualizations. The data was categorized into 4 categories: academic, social peer, extracurricular, and future. These categories are defined with a list of terms that relate to them (see table 9). Each dataset is scanned to keep track of the number of words identified from each category. A pie chart for each dataset inputted is generated, that represents these percentages, with their corresponding categories labeled. These visualizations illuminated the dominant stressors in adolescents' lives across platforms, offering a pathway to understanding how stress manifests and how interventions can address it.

6 Results

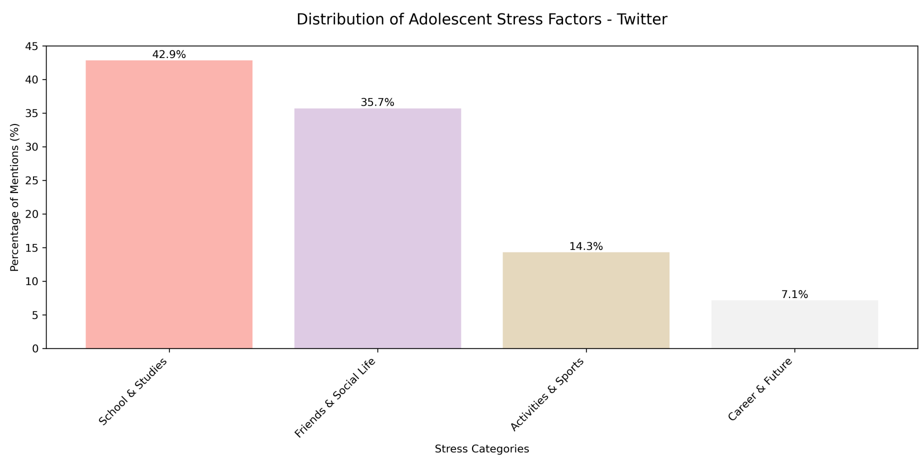


Fig. 1. Percentages of Negative Adolescent Stress expressed across Twitter

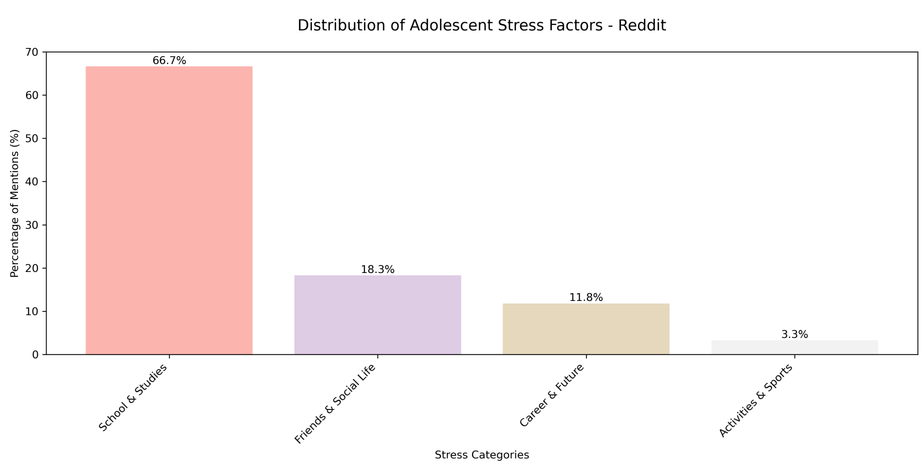


Fig. 2. Percentages of Negative Adolescent Stress expressed across Reddit

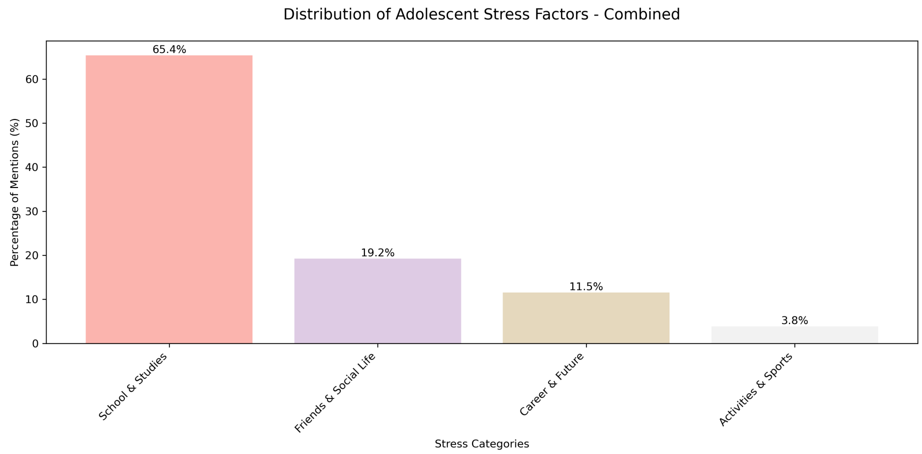


Fig. 3. Combined Percentages of Negative Adolescent Stress expressed on Reddit and Twitter

The histograms above illustrate the distribution of negative adolescent stress across four categories: School and Studies, Friends and Social Life, Activities and Sports, and Career and Future, analyzed separately for Twitter and Reddit and then combined.

A closer examination of the data reveals that, out of 51 rows in the combined Twitter datasets, 18 rows (35 percent) were classified as negative. For Reddit, 50 out of 225 rows (22 percent) were negative. These results highlight notable differences in the prevalence of negative stress between the two platforms.

Some of these difference may be explained by the accuracy of the LSTM model, which achieved approximately 50 percent for Twitter and 40 percent. The model's reliance on TextBlob for initial sentiment classification during training likely contributed to its limited performance. TextBlob, a basic classification tool, struggles to interpret social media-specific language, such as sarcasm and slang, which may have led to misclassification of sentiments.

Despite the limitations of the model, the histograms provide valuable insights into the primary sources of stress for adolescents on each platform. On Twitter, School and Studies emerges as the most prominent stressor, followed by Friends and Social Life, Activities and Sports, and Career and Future (see Figure 1). Similarly, on Reddit, School and Studies also leads, followed by Friends and Social Life, Career and Future, and Activities and Sports (see Figure 2).

When combining the results from both platforms, School and Studies stands out as the leading source of negative stress. In contrast, Activities and Sports contribute the least, likely because these are often perceived as leisure or hobby-related activities for adolescents (see Figure 3). These patterns reveal the importance of addressing academic pressures while recognizing the relatively smaller impact of extracurricular activities on negative stress levels.

7 Future Work/Discussion

Using this information, potential intervention strategies include targeted advertisements developed with mental health professionals to guide stressed adolescents to accessible support services. These services could include support groups addressing various topics, such as school or family, or providing contact information for helplines to offer immediate access to supportive conversations about stress-related challenges.

To maximize the impact of these ads, they could also display actionable strategies, such as engaging in physical activities or practicing mindfulness, along with their benefits, encouraging adolescents to take independent action. It is essential to ensure that ads promoting mental health support are as prevalent as those advertising other products. By maintaining a balanced presence, these ads can help normalize the pursuit of mental health support and make it easier for adolescents to identify and address their stress.

The necessity of these interventions becomes evident when considering the physiological impact of stress. Stress releases cortisol, a hormone responsible for regulating blood pressure and ensuring proper immune system function. However, excessive levels of stress can lead to overproduction of cortisol, impairing cognitive performance, causing high blood pressure, lowering immunity, and resulting in weight changes, insomnia, and lethargy ([17]). Therefore, maintaining regulated stress levels is critical for the optimal physical and mental health of adolescents and individuals of all ages.

The effectiveness of such targeted interventions is further supported by evidence from a study analyzing a digital advertising outreach campaign aimed at youth seeking mental health information in New York state. Within the first nine months, 55 teenagers and young adults were evaluated and referred to local care ([9]). This demonstrates how targeted ads can empower adolescents to address their stress and reduce the risk of developing serious health consequences.

8 Conclusion

In conclusion, this study utilizes LSTM models to analyze and classify negative stress expressed by adolescents on social media platforms, specifically Reddit and Twitter. The results revealed that academic pressures are the predominant source of stress among adolescents, as evidenced by the significant representation of school and study-related stressors in the analyzed datasets. By leveraging sentiment analysis, this research not only highlights the utility of deep learning techniques in uncovering stress patterns, but also provides actionable insights for designing targeted mental health interventions. Collaboration with mental health professionals can further enhance the effectiveness of strategies like digital campaigns and resource advertisements aimed at reducing adolescent stress. These findings underscore the potential of sentiment analysis to address critical societal challenges while enabling future research for further exploration into stress management solutions.

9 Acknowledgments

I would like to express my sincere gratitude to my professor, Dr. Neha Keshan, for her invaluable guidance and support throughout this research process. During our weekly meetings, she patiently reviewed my progress, provided constructive feedback, and offered insightful suggestions that significantly enhanced the quality of my work. Her encouragement and unwavering support were instrumental in helping me navigate challenges and achieve my goals.

Category	**Terms**
Stress Terms	stress, anxiety, depression, worried, nervous, panic, tension, pressure, overwhelm, burnout, mental health, emotional, exhausted, tired
Educational Terms	school, college, university, exam, test, homework, study, class, grade, assignment, teacher, professor, student, education
Age Terms	teen, teenage, adolescent
Slang Terms	lit, fam, salty, slay, tea, sus, ghosting, flex, shade, vibe, bet, lowkey, highkey, on fleek, woke, savage, bae, dope, snatched, no cap

Table 1. Categories and Terms

Layer Type	**Description**	**Purpose**
Embedding Layer	Converts each word into a dense vector of numbers.	Captures the meaning of the word in a lower-dimensional space.
LSTM Layer(s)	Long Short-Term Memory layers.	"Remembers" word patterns across the sequence to understand text flow and context.
Dense Layer	Fully connected layer with softmax activation.	Outputs the probability of each sentiment class (positive, neutral, negative).

Table 2. LSTM Model Structure for Sentiment Analysis

Category	**Terms**
Academic	exam, test, study, homework, grade, school, class, teacher, assignment, quiz, report, project, gpa, semester, final, subject, math, science, essay, tutor, academic, learning, education, classroom
Social Peer	friend, friends, party, hangout, social, peer, relationship, group, team, peer-pressure, best friend, drama, trust, support, loneliness
Extracurricular	sports, club, activity, extracurricular, hobby, interest, volunteer, practice, team, competition, performing, creativity, art, music, dance
Future Career	career, job, internship, future, resume, interview, experience, skills, networking, professional, education, goals, dream, ambition

Table 3. Categories and Relevant Terms

Dataset	**Purpose**	**Features (Columns)**
Reddit_Combi	Collection of combined title and body texts from Reddit posts, collected from both stress and non-stress related subreddits [3].	`title`, `body`, `body_title`, `label`
Reddit_Title	Contains titles from articles collected from both stress and non-stress related subreddits, aimed at stress classification tasks [3].	`title`, `label`
Mental-Health-Twitter	Targets mental health classification of users at the Tweet level, using various features of user engagement and sentiment [2].	`sentiment`, `post_id`, `post_creation`, `post_text`, `user_id`, `followers`, `friends`, `favorites`, `statuses`, `retweets`, `label`
Sentiment_140	Collection of tweets extracted using the Twitter API, used for general sentiment analysis [1].	`sentiment`, `id`, `timestamp`, `query`, `username`, `text`

Table 4. Datasets and Features

References

1. Dataset with 1.6 million tweets (September 2017), https://www.kaggle.com/datasets/kazanova/sentiment140, accessed on [Insert Date]
2. Twitter dataset + feature extraction (August 2022), https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media, accessed on [Insert Date]
3. Stress detection from social media articles (March 2024), https://www.kaggle.com/datasets/mexwell/stress-detection-from-social-media-articles, accessed on [Insert Date]
4. Al-Tameemi, I.K.S., Feizi-Derakhshi, M.R., Pashazadeh, S., Asadpour, M.: A comprehensive review of visual–textual sentiment analysis from social media networks. Journal of Computational Social Science (Sep 2024). https://doi.org/10.1007/s42001-024-00326-y, https://doi.org/10.1007/s42001-024-00326-y
5. Allen, S., Hiebert, B.: Stress and coping in adolescents. Canadian Journal of Counselling and Psychotherapy **25**(1) (May 2011), https://cjc-rcc.ucalgary.ca/article/view/59420
6. Ashok, P.: Urp 2024: Semantic analysis of adolescent stress. https://github.com/pashok4865/URP-2024-Semantic-Analysis-of-Adolescent-Stress (2024), gitHub repository
7. Asquith, N.: Understanding rnns, lstms and grus - towards data science. Medium (Mar 2022), https://towardsdatascience.com/understanding-rnns-lstms-and-grus-ed62eb584d90
8. Babu, N.V., Kanaga, E.G.M.: Sentiment analysis in social media data for depression detection using artificial intelligence: A review. SN Computer Science **3**(1) (Nov 2021). https://doi.org/10.1007/s42979-021-00958-1, https://doi.org/10.1007/s42979-021-00958-1
9. Birnbaum, M.L., Garrett, C., Baumel, A., Germano, N.T., Lee, C., Sosa, D., Ngo, H., Fox, K.H., Dixon, L., Kane, J.M.: Digital strategies to accelerate help-seeking in youth with psychiatric concerns in new york state. Frontiers in Psychiatry **13** (May 2022). https://doi.org/10.3389/fpsyt.2022.889602, https://doi.org/10.3389/fpsyt.2022.889602
10. Chauhan, G.S., Nahta, R., Meena, Y.K., Gopalani, D.: Aspect based sentiment analysis using deep learning approaches: A survey. Computer Science Review **49**, 100576 (2023). https://doi.org/https://doi.org/10.1016/j.cosrev.2023.100576, https://www.sciencedirect.com/science/article/pii/S1574013723000436
11. MentalHealth.com: Types of stressors (eustress vs. distress) (Sep 2024),
12. Molenaar, A., Lukose, D., Brennan, L., Jenkins, E.L., McCaffrey, T.A.: Using natural language processing to explore social media opinions on food security: sentiment analysis and topic modeling study. Journal of Medical Internet Research **26**, e47826 (Mar 2024). https://doi.org/10.2196/47826, https://www.jmir.org/2024/1/e47826/
13. Rastogi, A., Liu, Q., Cambria, E.: Stress detection from social media articles: New dataset benchmark and analytical study. In: 2022 International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2022). https://doi.org/10.1109/IJCNN55064.2022.9892889
14. Sharma, N.A., Ali, A.B.M.S., Kabir, M.A.: A review of sentiment analysis: tasks, applications, and deep learning techniques. International Journal of Data

Science and Analytics (Jul 2024). https://doi.org/10.1007/s41060-024-00594-x, https://doi.org/10.1007/s41060-024-00594-x
15. Team, H.E.: Common triggers of teen stress (2024), https://www.healthcentral.com/condition/anxiety/common-triggers-teen-stress, accessed: 2024-11-22
16. Whiting, A., Williams, D.: Why people use social media: a uses and gratifications approach. Qualitative Market Research an International Journal **16**(4), 362–369 (Aug 2013). https://doi.org/10.1108/qmr-06-2013-0041, https://doi.org/10.1108/qmr-06-2013-0041
17. Yassin, F.: The consequences of teenage stress (Jan 2023), https://thewaveclinic.com/blog/the-consequences-of-teenage-stress-2/