

Entrepreneurial Struggle: A Natural Language Processing Approach

Peter J. Yu

Louisiana School for Math, Science, and the Arts, 715 University Parkway, Natchitoches, Louisiana, 71457, USA; propeter35@gmail.com

ABSTRACT: Although entrepreneurial struggle is an important topic to explore in entrepreneurship research, it has received only limited attention from researchers. Some studies have begun to focus on the relationship between entrepreneurs' struggle and their mental health. However, there is little understanding of the issues that lead to entrepreneurial struggle. This research explores different areas in which entrepreneurs struggle in their venture-building process. Using a corpus of 10,150 semi-anonymous Reddit posts, this study identifies the areas of entrepreneurial struggle in the venture-building process. A total of 691 posts that contained struggle-related words were collected from the dataset. After applying Natural Language Processing (NLP) techniques, four areas of entrepreneurial struggle were found: product concept and business model; resources; market entry strategy and entry timing; and customer care, service, and communication. The findings of this study indicate that entrepreneurs should be prepared to deal with challenges in these areas in their entrepreneurial endeavors.

KEYWORDS: Behavioral and Social Sciences, Entrepreneur; Entrepreneurship; Entrepreneurial Struggle; Natural Language Processing; Machine Learning.

■ Introduction

While entrepreneurial activities are the engine of economic development¹, ninety percent of new ventures fail.² The high rate of new venture failure indicates that entrepreneurs need help in the venture-building process. In the entrepreneur's process of building new ventures, struggling with difficult tasks or situations is an important issue, but it has been understudied in contemporary research on entrepreneurship. Recent studies have started to explore entrepreneurial struggle^{3,4}, and more notably, the impact of struggle on mental health.⁵ However, despite the obvious importance of studying entrepreneurial struggle, the term has yet to be defined in the literature. In this paper, entrepreneurial struggle is conceptualized as the difficulty entrepreneurs encounter during the venture-building process.

It is critical to investigate struggle-eliciting events as such information would help inform entrepreneurs about potential setbacks they may encounter when trying to start a new business. If not dealt with successfully, those setbacks could lead to entrepreneurial failure, significantly influencing investors' support for the entrepreneur's subsequent ventures.⁴ Furthermore, venture failure has an impact on the confidence of entrepreneurs themselves.⁴ Confidence is a critical force that motivates the entrepreneur's venture-building effort and increases positive emotions while decreasing uncertainty.⁴ With no confidence, entrepreneurs who have struggled in the venture-building process may hesitate to try again.

Although entrepreneurial struggle may be a beneficial learning experience for entrepreneurs, it can be an emotionally negative and traumatic experience.⁶ Seventy-two percent of entrepreneurs reported mental health concerns including depression, ADHD, illegal substance abuse, and bipolar disorder.⁷ During the venture-building process, entrepreneurs often work in an unpredictable environment that requires many tasks

for which they are frequently unprepared. In addition, mental health for entrepreneurs is critically important.^{8,9} To improve our understanding of mental health in entrepreneurship, it is crucial to identify and study struggle-eliciting events in the venture-building process.

As entrepreneurial struggle can lead to business failure and mental health issues, it is a highly relevant topic in entrepreneurship scholarship. However, the limited research on the aspects of entrepreneurial struggle has been mostly *post hoc* and has yet to explore the struggle-causing events as they happen.^{6,10} Furthermore, the types and nature of entrepreneurial struggle have not been studied in depth. The current study investigates the areas in which entrepreneurs struggle and aims to achieve a better understanding of entrepreneurial struggle across a wide range of business contexts.

■ Literature Review

Researchers have realized the importance of studying entrepreneurial struggle. For example, past research has examined the relationship between mental health and entrepreneurial failure. Often, mental health issues are a result of struggling through the process of starting a new business. Researchers explored the roles of confidence and emotion in the venture-building process. Failed ventures, which are associated with struggle-related events, can affect entrepreneurs' emotional and cognitive resilience and confidence.⁴ They propose five outcomes that are more likely to exist among confident entrepreneurs than those less confident: positive emotions, emotional resilience, social support from team members, financial resilience, and subsequent ventures. In addition, greater failures in venture building can damage entrepreneurs' reputations, which can critically impact social and financial capital.⁴

Other researchers investigated the role of depression in entrepreneurial exit. Depression impacts an entrepreneur's personality, which plays a role in the intention to start a new

business and business success.⁹ Furthermore, depression has an impact on self-efficacy. Entrepreneurs with low self-efficacy may perceive complex tasks as threats and negatively evaluate themselves.⁹ Low self-efficacy increases the likelihood of exiting the venture-building process. On the other hand, entrepreneurs with high self-efficacy will commit more and persevere in challenging tasks.⁹ Out of the 12,293 cases the researchers investigated, there are 2,496 instances of entrepreneurial exit.⁹ The researchers found that depression positively correlates with exiting from self-employment, and self-efficacy mediates 32% of the relationship between depression and exit.⁹

Recently, several studies have attempted to use machine learning and Natural Language Processing (NLP) techniques to investigate entrepreneurial mental health using social media.^{5,11-13} For instance, one study used machine learning on a corpus of 27,906 semi-anonymous posts on Reddit to study entrepreneurial disappointment.⁵ A supervised machine learning classification task was performed, and the Logit Boost algorithm was selected to detect disappointment-related posts with an 88% accuracy.⁵ The study established five attributions of entrepreneurial disappointment: self-related, norms-related, others-related, entrepreneurship-process-related, and venture-performance-related.⁵ It found that internal, global, and stable causes resulted in a higher frequency of depression compared with external, specific, and temporary causes.⁵ Although this study provided valuable insight regarding the association between disappointment and entrepreneurship, the factors that lead to entrepreneurs' disappointment were not investigated.

Using NLP techniques to analyze social media posts on platforms such as Reddit has also been carried out by other researchers. These studies demonstrated the value of this particular research methodology. For example, researchers from a recent study collected a corpus of 22,808 posts on Reddit over a period of 3 months to study anxiety disorders.¹² To classify anxiety-related posts, they generated features of the posts by applying Natural Language Processing techniques, more specifically N-gram language modeling, vector embeddings, topic analysis, and emotional norms.¹² They were able to classify posts with 98% accuracy.¹² Latent Dirichlet Allocation (LDA) topic modeling, an unsupervised feature generation technique, was used to find correlations between specific topics and anxiety.¹² This method could be very effective in identifying struggle-related topics that entrepreneurs run into during the venture-building process.

Another study employed machine learning and NLP techniques to detect the presence of depression in Reddit posts. Specifically, it used a MultiLayer Perception (MLP) classifier, a combination of Linguistic Inquiry and Word Count (LIWC), LDA, and bigram techniques to achieve a 91% accuracy in detecting depression in Reddit posts.¹³ The researchers discovered that depression-related posts contained words connected to anxiety, sadness, and other negative emotional states.¹³ Although their research successfully detected depression in Reddit posts, it did not investigate the reasons behind the negative emotional states.

In short, social media has become a popular resource for detecting and predicting mental issues and other topics. These

studies have used artificial intelligence and computational linguistics to utilize social media datasets.^{5,12,13} This paper takes inspiration from them to identify the types and nature of entrepreneurial struggle.

■ Methods

Data Collection:

Reddit is one of the largest social networks for news aggregation, content rating, and discussion. Currently, Reddit has over 50 million daily active users and 100,000 communities.¹⁴ The platform allows posts to have relatively large bodies of text compared with other social networks. In addition, unlike other social media sites, Reddit offers anonymity to users. This enables candid discussions of the problematic issues, and hardships entrepreneurs deal with.

The "r/startups" subreddit comprises more than one million members who discuss starting new ventures. A total of 10,150 posts were extracted from the "r/startups" subreddit group from January 1, 2020, to October 1, 2022, using the Pushshift API.¹⁵ To search for posts related to entrepreneurial struggle, a word list containing struggle-related words was used to filter through the corpus. The word list is given below:

anxiety, worry, fret, concern, agitation, apprehension, nervous, unease, stress, pressure, trouble, struggle

It is important to note that the algorithm used in the study included all the verb tenses where appropriate, which are not shown above. Many of the posts confirmed the importance of user anonymity in this study. For example, one anonymous entrepreneur stated: "I have been dealing with many doubts and am unmotivated. I am thinking about leaving my own startup. I need support from someone in the startup world, and everyone I confront doesn't fully understand the problem." Posts like this were viewed and discussed by many Reddit users.

Of the original corpus, 691 posts contained terms related to entrepreneurial struggle. A series of NLP tools were performed to pre-process the dataset. First, URLs, punctuation, and stop words such as "the," "a," and "I" were removed from the dataset through the Natural Language Toolkit¹⁶ library in Python because they do not contribute to our ability to understand the substantive issues under study. Next, lemmatization was performed on each post to remove inflectional endings or to return the base forms of words. Finally, lemmatization was used instead of stemming because the context of the word is considered.

Topic Modeling:

Topic modeling is an unsupervised statistical modeling technique that identifies topics that describes a set of documents. In this research, Latent Dirichlet Allocation (LDA), a generative topic modeling technique, is used to classify text in the posts into specific topics.¹⁹ To use the LDA topic modeling approach, the study has to specify the number of topics in the model. An LDA model was trained on the dataset to generate latent unlabelled topics characterized by a selected distribution of the top 10 individual words.

Following the data pre-processing, Term Frequency-Inverse Document Frequency (TF-IDF) vectorization was used to transform text into numerical data and highlight the importan-

ce of words in the corpus. TF-IDF is a statistical measure calculated by multiplying the term frequency and inverse document frequency.¹⁷ The term frequency can be found by the raw count of a word that appears in a document. Term Frequency (TF) is captured in equation 1, where t represents a term and d is a document.

$$TF(t, d) = \frac{\text{number of times } t \text{ appears in } d}{\text{total number of terms in } d} \quad (\text{Eq. 1})^{18}$$

Inverse Document Frequency (IDF) is a statistical measure of how common or rare a word is across a document corpus. The IDF can be calculated by taking the logarithm of the total number of documents and the number of documents that contain the term.¹⁷ To avoid division by zero, one is added to the number of documents that contain the term. In equation 2, IDF is captured, where N is the total number of documents and df is the number of documents that contain the term t .

$$IDF(t) = \log \frac{N}{1 + df} \quad (\text{Eq. 2})^{18}$$

TF and IDF are then multiplied to find the TF-IDF, represented in equation 3.

$$TF - IDF(t, d) = TF(t, d) * IDF(t) \quad (\text{Eq. 3})^{18}$$

The proposed number of topics in the LDA model was adjusted several times and some of them had unbalanced word distributions and needed to display clear themes. For example, when the proposed number of topics was 6, posts in one topic had keywords that were not closely related to one another, such as market timing and product conceptualization. After experimenting with models that have 3 – 7 topics, the 4-topic solution was deemed the best as all topics had relatively consistent themes.

■ Results and Discussion

Latent Dirichlet Allocation Topic Modeling Results and Discussion:

The results of the LDA model when the proposed number of topics was four are shown in Table 1. Most of the top topic words for Topic A are related to finding a product concept or business model. In other words, posts assigned to this topic indicate entrepreneurs needed help understanding the market, the needs of customers, and the product or service they should offer. For Topic B, the top topic words primarily hinted at a struggle for resources. These words show that entrepreneurs in this category are looking for solutions regarding funding, the time needed for the business, human resources, etc. Topic C has top words related to the struggle of deciding the best market entry strategy and entry timing (e.g., plan, time, and start). The “time” topic word for Topic C differs from the one in Topic B because it relates to the optimal timing of marketing actions such as when to enter the market. These entrepreneurs seem to need help coming up with a plan to enter the market with the right approach for the target market. Finally, the top words for Topic D are associated with customer care, service, and communication.

Table 1: Top topic words.

	Top 10 Topic Words
Topic A	market, business, people, customer, use, idea, build, user, app, product
Topic B	start, investor, job, cofounder, make, time, year, work, startup, company
Topic C	talk, plan, year, need, people, time, work, make, start, business
Topic D	person, small, provide, try, solution, platform, website, appreciate, create, sell

Of the 691 posts containing struggle-related words, 321, 251, 102, and 17 were assigned to Topic A, B, C, and D, respectively. The distribution of each struggle category can be seen in Figure 1.

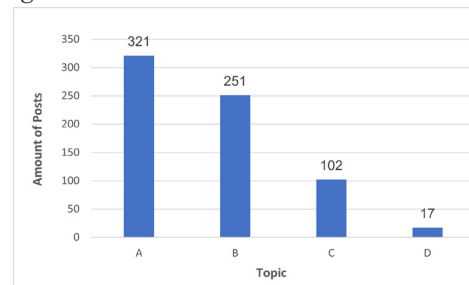


Figure 1: Distribution of posts among topics.

Sample Posts:

To help illustrate how entrepreneurs discuss their struggles in the four topics, this section provides some sample posts from the four topics. As discussed earlier, Topic A is about entrepreneurial struggle regarding the business model or product idea. This is evident in the post below:

“Sometimes life can be hard or too much, and you need someone who isn't a part of your life to talk to. Therapists, friends, and family want to fix you. But, sometimes, finding someone who wants to listen to or talk with you is hard. I really want to start a business that helps people and gives them someone to talk to in these times. It helps those with social anxiety to have a friend. I want to offer some people free trials in the beginning because I want to know what they would want that would make it better for them. So, if anyone is interested or you know anyone that would be interested let me know. To be clear, I am not a therapist or a mental health professional and am just wanting to start a business for people to be able to pay for convenient friendship.”

In the post above, the would-be entrepreneur had a vague idea of a potential product for their intended market but needed to know how attractive the product would be. Thus, they reached out to the audience on Reddit to discuss. This person apparently has a strong desire to understand the target market and tries to provide a service that could meet the market needs. This pattern could be observed in many posts in the Topic A category.

Posts in the Topic B category reported resource struggles (e.g., money, time, and human resources). These entrepreneurs are putting in the effort to assemble the necessary resources to build their businesses but seem to need more of them. For example, in the sample post below, one entrepreneur had difficulty in preparing for pitching the business to potential investors:

“Hi everyone! We’re a toy startup preparing for our first investment round. Honestly, my team and I are nervous about it,

and I wanted to ask for advice from people with actual pitching experience. How did you get in touch with funds? Did you have a warm intro? What's the most important in your opinion?"

Market entry strategy and timing were discussed by entrepreneurs who requested help with posts in Topic C. These entrepreneurs need to figure out the optimal market strategy, including market timing. For example, the person that posted the request below was trying to figure out the appropriate price for a new app, with consideration for the affordability of the product in different geographical locations.

"I have recently launched an app and have struggled with finding the right price to charge the users. I know that people from the USA and Europe can pay higher prices, but I feared that it was too expensive for the rest of the world. I looked into the ability to charge different prices based on users' country and could not find a product like that other than for WooCommerce. My website is built from the ground up so that is not an option. My question to ask is whether you guys already do this. Do you know of a web app that does it? Would you use something like this?"

The final topic area, Topic D, includes posts regarding providing the best customer care, service, and communication. One entrepreneur, who had the goal of targeting senior customers, struggled with communicating with the targeted group:

"Hi, all. About a month ago, I launched a product for seniors and retired people to get 100 users in 100 days. Are there any suggestions on how I could get those 100 users? I have two personas: people aged 65+ with specific habits and their 35–45-year-old children, who can also sign up for the service for their parents. To test product-market fit, I'm trying to test a few channels: google ads (10% CTR), Facebook groups (very positive comments, dozens of clicks, but no traction), forums, and emails. I managed to scrape hundreds of emails from several communities and launched a cold email campaign. There was a 60% open rate and many clicks, but no results. I also initiated some partnerships with websites with a senior audience in order to get referral traffic: crosswords, sudoku, knitting, and gardening. A few answered back and one of them accepted. Long story short, I'm struggling to find potential clients for the service. Facebook ads are an option of course, which I will test, but my budget is not unlimited. Any ideas that could help? Thank you so much for your support!"

■ Discussion

While previous research has investigated some negative aspects of the venture-building process such as entrepreneurial anxiety, there is little research on the types and nature of entrepreneurial struggle. This research suggests that entrepreneurs need help in four primary areas: product concept and business model; resources; market entry strategy and entry timing; and customer care, service, and communication. Most of the struggles are in the marketing area. This points to the difficulty entrepreneurs often have in connecting what they contemplate offering to the market and the market itself. Past research has not sufficiently investigated entrepreneurs' struggles in the marketing area and the constraints of limited resources. Topics

such as product model, resources, market strategy, and customer service are explored for the first time in the current study. It could be argued that marketing and assembling resources to make the new venture a reality are the two main factors hindering entrepreneurs' efforts to become successful.

Based on these findings, entrepreneurship education programs should focus on the four struggle areas identified in this study to prepare entrepreneurs better before starting ventures and during the venture-building process. Teaching entrepreneurs in areas such as marketing and fundraising is most needed. This will help entrepreneurs acquire knowledge and skills critical to starting companies and improve the likelihood of new business success. On the other hand, new business owners and aspiring entrepreneurs should understand that they must develop skills in the four areas identified.

Although previous studies have, to some extent, researched the affective aspect of entrepreneurial struggle^{5,8}, the term "entrepreneurial struggle" until now has been used without an explicit definition. In this paper, we define entrepreneurial struggle as the difficulty entrepreneurs encounter during the venture-building process. This definition allows scholars to distinguish between entrepreneurial struggle and events that may not be related to the entrepreneurial process, such as personal or family issues. Furthermore, more clarity is established for future entrepreneurship research by defining entrepreneurial struggle this way.

Intense mental illness symptoms can arise because of struggle-related events entrepreneurs face.⁵ The resulting emotional response after entrepreneurs face challenges may lead to depression and unproductive behavioral changes. Research has found that entrepreneurs may stop socializing and isolate themselves when coping with negative emotional responses.⁵ Conversely, entrepreneurs who expressed disappointment because of loneliness had the desire to maintain meaningful relationships. The behavioral response in entrepreneurs, when faced with struggle-related events, is not beneficial for their mental health.⁵ This further points to the importance of helping entrepreneurs deal with the struggle areas identified in this research as doing so will have a direct positive impact on the mental health of entrepreneurs. For example, if given the opportunity, an entrepreneur who has never had experience in communicating with customers could improve the skills required in this area and have a better experience in the entrepreneurial journey. Therefore, this enhancement in skills would lead to an improvement in the entrepreneur's mental health.

Limitations and Future Research:

The findings and limitations of this study have important implications for future research in entrepreneurship. Although several types of entrepreneurial struggle were identified in this study and their relations with the affective experience of entrepreneurs pointed out, it still needs to be made clear whether entrepreneurs with mental health issues encounter more struggle during the venture-building process. In other words, it remains to be investigated whether mental health issues and entrepreneurial struggles actually influence each other. Furthermore, while the results of this study show that entrepreneurs primarily struggle in the four areas, the extent of

the effect of the struggle-related events on new venture success should have been researched. Further research is required to investigate this issue as some areas of struggle may lead to new business failure more than others.

While Reddit provides anonymous posting and allows for researching topics associated with a stigma or shame for those posting messages, it does not let researchers examine the entrepreneurs' previous venture experiences, industries, and other backgrounds.⁵ These factors may directly impact the extent to which entrepreneurs experience potential struggles and how they deal with struggles. When posts were extracted from Reddit, only the body text of posts was collected. Any information that could reveal the post's author (i.e., username) was not collected. The identities and descriptions of the entrepreneurs were kept unknown throughout the study.

Furthermore, the Reddit user base is different from the general population. According to Pew Research Center, the Reddit user base in the United States is majority male (67% of users), white (70% of users), and young (64% of users are under 29 years of age).²⁰ Additionally, Reddit users have higher levels of digital literacy and web access.⁵ Therefore, this study may have yet to capture entrepreneurs with different characteristics.

Also, the mood and circumstances under which the entrepreneurs made the posts could influence the contents of the posts themselves. For example, it is possible that some entrepreneurs posted during a night of poor sleep or a cold. Future studies should make efforts to identify these conditions and their associated effects. Additionally, when a team of entrepreneurs starts new ventures, more than one individual entrepreneur can encounter difficulty building a new venture. The interactions and relationships among a team of entrepreneurs may influence decisions when building new ventures. These decisions may be crucial when acquiring resources and meeting public expectations. Exploring struggle among a team of entrepreneurs is a valuable extension of entrepreneurial struggle research.

Finally, future research can examine the effects of entrepreneurial struggle on the personal lives of entrepreneurs. Although the struggle areas identified in this study impact the venture-building process, more investigation is needed to examine the relationship between struggle and entrepreneurs' life and career paths. In addition, further research is required to explore how entrepreneurs constructively grow and build resilience following entrepreneurial struggle and their inclination to continue being an entrepreneur.

■ Conclusion

In this study, 10,150 semi-anonymous posts were extracted from Reddit to examine the struggles of entrepreneurs. After implementing NLP techniques, four primary struggle areas that entrepreneurs encountered were identified: product concept and business model; resources; market entry strategy and entry timing; and customer care, service, and communication. As entrepreneurial struggle is connected with entrepreneurial failure and the mental health of entrepreneurs, there are numerous directions for future research in this area. It is also essential to use the findings of this research to create educational and training materials to help entrepreneurs better prepare for the

inevitable trials and tribulations during their venture-building endeavors. Organizations, government agencies, and educational institutions that provide entrepreneurship education and assistance programs should especially pay attention to the four struggle areas identified in this paper in their programs to help entrepreneurs cope with the challenges they encounter.

■ Acknowledgments

I want to acknowledge the mentorship of Professor Brad Burkman of the Louisiana School for Math, Science, and the Arts. Furthermore, I am grateful to Dr. Lidan Fan of the University of Kansas, Dr. Long Pham of Texas A&M University, and Dr. Javier Flores of Emporia State University for their comments that helped improve this paper. In addition, I would like to thank my instructors at the Louisiana School of Math, Science, and the Arts: Dr. Zheng Chen for motivating me to grow my knowledge in STEM, Professors Randy Key and David Andersen for inspiring my interest in mathematics, and Professor Sanjeetha Peters for encouraging me to pursue computer science. Finally, I am grateful to my parents for their support.

■ References

1. Meyer, N.; de Jongh, J. The Importance of Entrepreneurship as a Contributing Factor to Economic Growth and Development: The Case of Selected European Countries. *JEBs* **2018**, *10*, 287–299.
2. Patel, Neil. 90% of Startups fail: Here's What You Need to Know About The 10%. *Forbes* 2018. <https://www.forbes.com/sites/neilpatel/2015/01/16/90-of-startups-will-fail-heres-what-you-need-to-know-about-the-10/?sh=4524c8926679> (accessed November 15, 2022).
3. DeTienne, D. R. Entrepreneurial exit as a critical component of the entrepreneurial process: Theoretical development. *Journal of Business Venturing* **2010**, *25*(2): 203–215.
4. Hayward, M.L.A.; Forster, W.R.; Sarasvathy, S.D.; Fredrickson, B. L. Beyond hubris: How highly confident entrepreneurs rebound to venture again. *Journal of Business Venturing* **2010**, *25*(6): 569–578.
5. Williamson, A. J.; Drencheva, A.; Battisti, M. Entrepreneurial Disappointment: Let Down and Breaking Down, a Machine-Learning Study. *Entrepreneurship Theory and Practice* **2022**, *46*(6), 1500–1533.
6. Ucbasaran, D.; Shepherd, D. A.; Lockett, A.; Lyon, S. J. Life After Business Failure: The Process and Consequences of Business Failure for Entrepreneurs. *Journal of Management* **2013**, *39*(1), 163–202.
7. Freeman, M. A.; Johnson, S. L.; Staudenmaier P. J.; Zisser M. R. Are Entrepreneurs "Touched with Fire?" Unpublished Manuscript 2015 (accessed November 16, 2022).
8. Baron, R. A. The Role of Affect in the Entrepreneurial Process. *The Academy of Management Review* **2008**, *33*(2), 328–340.
9. Hessels J.; Rietveld C.A.; Thurik A.R.; Zwan P.W. Depression and Entrepreneurial Exit. *Academy of Management perspectives* **2018**, *32*(3): 323–339.
10. Thompson, N. A.; Gelderen M. V. Coping with anxiety entrepreneurial process. *Academy of Management Proceedings* **2018**; Academy of Management: Briarcliff Manor, New York.
11. De Choudhury, M.; Gamon, M.; Counts, S.; Horvitz, E. Predicting Depression via Social Media. *ICWSM* **2021**, *7*, 128–137.
12. Hanwen Shen, J.; Rudzicz F. Detecting Anxiety on Reddit. In Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality 2017: Vancouver, BC, 58–65.
13. Tadesse M. M.; Lin, H.; Xu B.; Yang L. Detection of Depression-Related Posts in Reddit Social Media Forum. *IEEE Access* **2019**, vol. 7, pp. 44883–44893.

14. Proferes, N.; Jones, N.; Gilbert, S.; Fiesler, C.; Zimmer, M. Studying Reddit: A Systematic Overview of Disciplines, Approaches, Methods, and Ethics. *Social Media + Society* **2021**, 7(2).
15. Baumgartner J.; Zannettou S.; Keegan B.; M. Squire; Blackburn J. The Pushshift Reddit dataset. In *Proceedings of the International AAAI Conference on Web and Social Media 2020*, vol. 14, pp. 830-839.
16. Loper E.; Bird S. NLTK: The Natural Language Toolkit, 2002. arXiv.org e-print archive. <https://arxiv.org/pdf/cs/0205028.pdf> (accessed January 29, 2023).
17. Qaiser, S. & Ali, R. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications* **2018**.
18. Chen K. Introduction to Natural Language Processing – TF-IDF, 2021. Medium. <https://kinder-chen.medium.com/introduction-to-natural-language-processing-tf-idf-1507e907c19> (accessed November 17, 2022).
19. Blei, D. M., Ng, A. Y.; Jordan, M. I. Latent Dirichlet Allocation. *Journal of machine Learning research* **2003**, pp. 993-1022.
20. Barthel, M.; Stocking, G.; Holcomb, J.; Mitchell, A. Nearly eight-in-ten Reddit users get news on the site. Pew Research Center 2016.

■ Author

Peter Yu is a sophomore at the Louisiana School for Math, Science, and the Arts in Natchitoches, Louisiana. He participates in competitive swimming and plays the piano. In the future, he plans to pursue computer science in college.