

NLP to Partition Reddit Posts Related to Neurodiversity and Employment

Michael West
Department of Computer and Information Sciences
Temple University
Philadelphia, PA USA
michael.west0004@temple.edu

Maddie Upthegrove
Department of Psychology and Neuroscience
Temple University
Philadelphia, PA USA
mupthegrove@temple.edu

Abstract

The present project utilized topic modeling and document clustering to partition a text corpus from r/autism and conducted a mixed methods analysis of the resulting document subsets. We compare the results of two approaches, topic modeling with LDA and document clustering with K-Means clustering applied to a tf-idf document representation. Findings show that tf-idf was superior to LDA, but suggest that there may be underlying interactions between the groupings of additional posts of interest which presently elude each model separately. When groupings of each model are applied, the larger dataset is able to successfully be partitioned for further analysis of employment. A hand review of these overlapping posts shows promising results for the validity of these topic discovery methods. Further analysis of these posts can provide researchers and practitioners with unique insights into the experiences, challenges and barriers surrounding employment for adults with ASD.

Presenting Problem

The nature of work has shifted increasingly towards computer-based work which has resulted in widespread labor shortages as markets become more and more competitive [1]. While these shortages are severe, [2] suggests that there are groups of the population which have historically been weeded out of the population of able workers through the hiring process. One of these groups are neurodiverse workers who have been found, when given adequate supports and accommodations, to outperform their neurotypical peers [3]. Neurodiverse employees have strengths such as concentration and patience for repetitive work, excellent accuracy, reliability and dependability, and trustworthiness, honesty, and integrity [4].

While neurodiversity is a broad category, autism is one of the most prevalent diagnoses with diagnostic criteria [5] including difficulty with communication and interaction with other people, restricted interests and repetitive behaviors, and symptoms that affect their ability to function in school, work, and other areas of life. Autistic workers commonly have challenges adapting to changes in routine as well as social difficulties in the workplace [4]. While these challenges are most successfully addressed with employer involvement, many employees and prospective employees are left to address them on their own.

Reddit and r/autism

Numerous resources exist for people to seek “self-help,” one popular source is a network of web-based online forums, or communities, named Reddit. Reddit has individual forums, or subreddits, devoted to specific topics in which users can submit and respond to media or text content. Each sub-reddit is a unique group which may have its own culture including terminology, common practices, and moderators who monitor the subreddit to ensure compliance with rules.

One subreddit of particular interest to the issue of underemployment of neurodiverse workers is the subreddit called “autism,” or r/autism. This subreddit provides not only a community oriented towards autism, but also an opportunity for internet-based communication with a more reduced social expectations than in person communication. Due to the prevalence of difficulties with social interactions, this presents an opportunity for autistic

employees and prospective employees to seek assistance with job related questions in a safe and secure setting with fewer social barriers.

Purpose

Compare two topic discovery methods, latent dirichlet allocation (LDA) and document clustering, to partition the larger dataset that is specifically related to work for further mixed methods analysis.

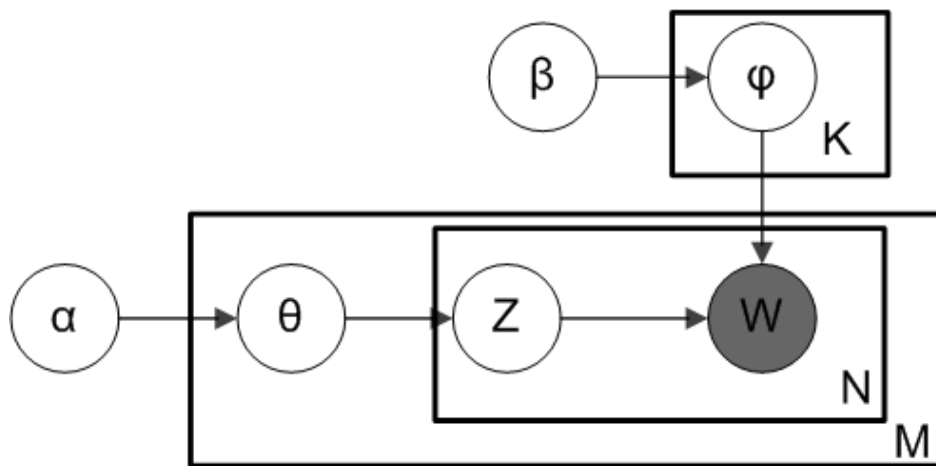
Topic Discovery

Topic Discovery is a subproblem of Natural Language Processing that seeks to group documents from a corpus into subsets such that all documents in a subset discuss the same semantic topic. It is analogous to the general machine learning problem of clustering. Much like clustering, topic discovery models are evaluated on their ability to maintain within group similarity while creating external differentiation. In this paper we utilize two methods for topic discovery, Latent Dirichlet Allocation (LDA) and document clustering implemented with term-frequency inverse document-frequency (tf-idf) representation with K-Means clustering.

LDA

LDA is an unsupervised generative statistical model commonly used for topic discovery. LDA posits that each document in a corpus is a combination of underlying semantic topics and that the presence of each word in a document is due to one of the underlying topics. LDA seeks to discover the underlying semantic topics in a corpus. For each discovered topic, the model assigns each document in the corpus a score indicative of how related that document is to the given topic.

The underlying model for LDA is shown in the figure below. The boxes represent repeated entities. Box M represents the documents of the corpus, box N represents the words in a document, and box K represents the topics. α and β are model parameters representing the Dirichlet prior on per document topic distributions and per topic word distributions respectively. θ is the topic distribution for the given document and ϕ is the word distribution for the given topic. Finally, W is the given word and Z is the topic associated with that word. To infer the topics in a corpus, LDA assumes that this underlying model is used in a generative process to produce the documents, and then reverse engineers that process to discover the underlying topics.



One limitation of using LDA for topic discovery is the difficulty for humans to interpret and draw conclusions from the topics generated [6]. A major contributing factor to this issue is the difficulty in visualization of LDA outputs. LDAvis attempts to address these difficulties by creating a web-based interactive visualization of topics that presents the topics using a visualization system that presents the topic model globally as well as presents terms ranked according to relevance, rather than probability alone. Relevance is calculated as follows:

$$r(w, k | \lambda) = \lambda \log(\phi_{kw}) + (1 - \lambda) \log\left(\frac{\phi_{kw}}{p_w}\right)$$

Where lambda is the weight given to the probability of term w under topic k relative to its lift where lift is the ratio of a term's probability within a topic to its marginal probability across the corpus. While a lambda equal to one would return the probability alone and a lambda equal to zero would return the lift alone, the recommended lambda value is .6 [6].

tf-idf

As an alternative to topic modeling, document clustering can be applied for topic discovery. In this approach documents are transformed into vector representations and then a traditional machine learning algorithm for clustering, such as K-Means clustering is applied to the vector representations. We implemented this approach with K-Means clustering applied to a tf-idf vector representation. Tf-idf is a term weighting scheme wherein the frequency of appearance of a term in a given document is offset by the frequency of appearance of the term in all documents in the corpus. For a given term t , in document d , the tf-idf is given by the equation:

$$tfidf = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \cdot \log \frac{N}{d \in D: t \in d}$$

where $f_{t,d}$ is the frequency of term t in document d and N is the total number of documents in the corpus. The result of a tf-idf embedding is a sparse vector representation for each document in the corpus.

Similar Work

Topic discovery is a well-studied problem in the field of NLP. Techniques vary in implementation, but generally they involve transforming text documents to a vector representation, often referred to as a document embedding. Once documents are mapped onto a vector space, machine learning algorithms can be applied to partition the documents into topic specific subsets. These techniques have seen some success in topic discovery for well-structured text documents such as books, research papers or news articles [7]. However, there are issues when applying these techniques to content from online platforms, as these documents are comparatively noisy [8].

Previous research has investigated topic discovery for data obtained from reddit. One paper compared document clustering and topic modeling techniques for data obtained from multiple subreddits [9]. However, this work used the subreddit of origin as a proxy for the topic of the document and aimed to classify new documents into one of these topics. Another paper aimed to discover a specific topic (electronic cigarettes) in reddit data [10]. However, this work conducts the topic discovery across multiple different subreddits. The present work varies in that we focus our analysis on a specific community, r/autism, and seek to discover a specific topic, employment.

Additionally, recent research has studied data from r/autism. One study analyzed the sentiment of discussions related to an experimental intervention for autism, Applied Behavioral Analysis [11]. Another study conducted a qualitative analysis of posts seeking support for siblings of those with autism [12]. The present work varies in methodology, as we combine quantitative and qualitative analysis of topic discovery.

Methods

Dataset

The Reddit Mental Health Dataset [13] contains posts gathered between 2018 and 2020 from 15 specific mental health related subreddits, 2 broad mental health subreddits, and 11 control subreddits from 826,961 unique users. The authors of this dataset conducted a broad analysis comparing subreddits in the mental health group to subreddits in the control group. Their analysis largely focused on changes in mental health in relation to the COVID-19 pandemic.

Preprocessing

From the Reddit Mental Health Dataset [13], we extracted all posts from r/autism, yielding 8869 unique posts from 7177 unique users. A number of pre-processing steps standard to NLP were applied to this data. First, all text was normalized to lower case and punctuation was removed. Additionally, reddit allows users to utilize markdown formatting, therefore markdown specific special characters were removed. Finally, stop words were removed using the English stop word list from the popular python package *nlk*.

LDA

We implemented LDA using the popular python package *gensim*, and a chosen number of topics 10. Another python package, *pyLDAvis*, was used to visualize and manually analyze the results. Additionally, documents were assigned to their most relevant topic and the most frequent terms and phrases from each topic group were manually analyzed. Phrases are represented by 3, 4 and 5 grams, with the inclusion of stop words. Lambda = .6 as recommended by [6], was used for visualization and qualitative analysis.

Tf-idf

We implemented tf-idf with the python package *sklearn*. Tf-idf includes two hyperparameters, maximum document frequency and N-gram range, Maximum document frequency sets a threshold, in the range 0 to 1, of document appearance for which any term above this threshold is excluded, such that 0 will exclude every term and 1 will exclude no terms. This essentially creates corpus specific stop-words. Values of 0.25, 0.50, 0.75 and 1.0 were tested. N-gram range sets a range of n-grams to be considered as terms such that (1,1) utilizes only unigrams, (1, 2) utilizes both unigrams and bigrams, and so on. Values of (1,1), (1,2) and (2,2) were tested. Larger ranges were not tested due to computational limitations arising from combinatorial explosion when attempting to include longer n-grams as terms. Once the tf-idf representations were obtained, K-Means clustering was applied and resulting clusters were evaluated with silhouette score. Silhouette score is a common internal validation metric for K-Means clustering which measures the balance of within cluster similarity and external cluster differentiation using a Euclidean distance metric. The number of clusters, k , was tested with values between 2 and 10 inclusive. After clustering, frequent term and phrase analysis was conducted.

Results

LDA Results

Results were posts sorted into ten topics and multidimensional scaling was used to create an intertopic distance map. Size of topics varied from 58 to 2451 with a mean of 886.5 topics and a standard deviation of 882.0 topics.

The top-30 most relevant terms were also generated with lambda equal to 0.6, for the corpus as well as for individual topics. The grey bars represent the global frequency of each term and the widths of the red bars represent the topic-specific frequencies of each term.

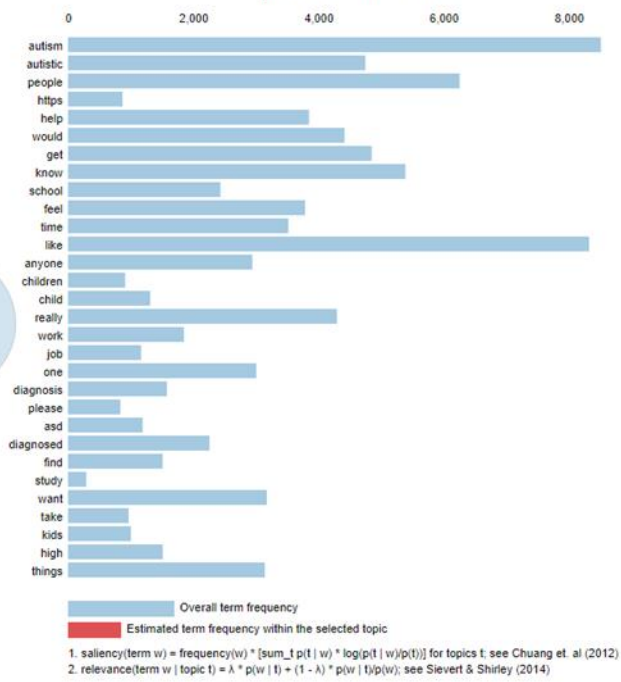
Selected Topic:

Slide to adjust relevance metric:⁽²⁾ $\lambda = 0.6$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Salient Terms¹

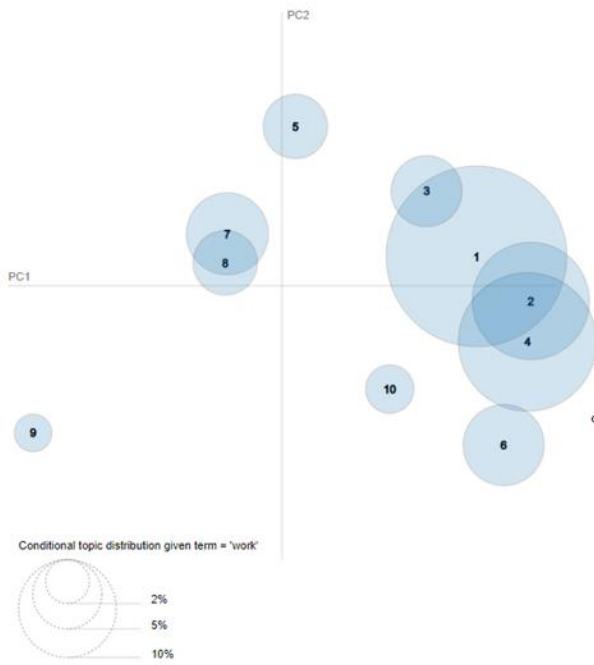


The terms related to the presenting issue of work that were present in the top 30 most salient terms were work and job. Their prevalence in the ten topics are presented below:

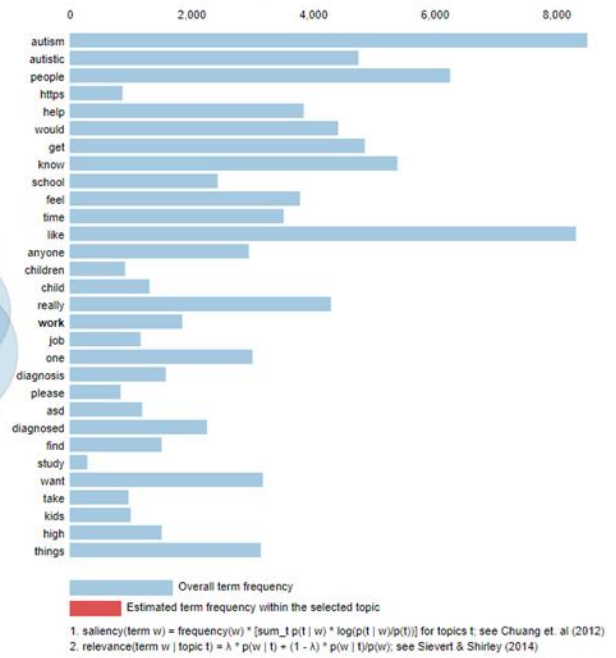
Selected Topic:

Slide to adjust relevance metric:⁽²⁾
 $\lambda = 0.6$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Salient Terms¹



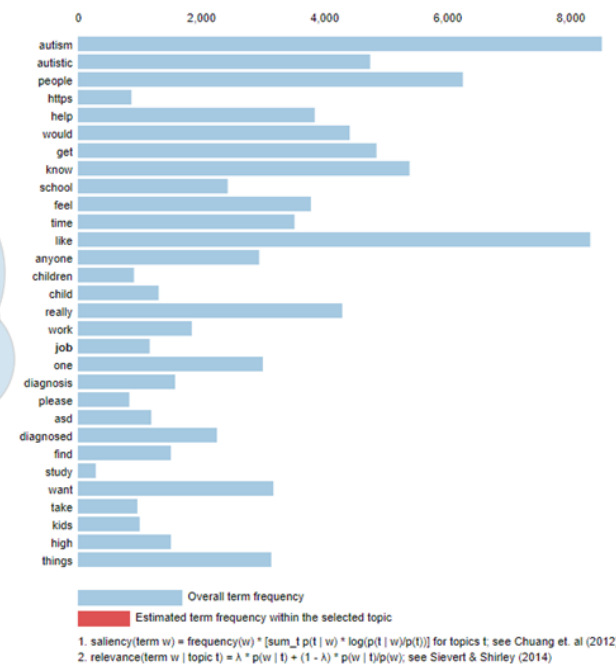
Selected Topic:

Slide to adjust relevance metric:⁽²⁾
 $\lambda = 0.6$

Intertopic Distance Map (via multidimensional scaling)



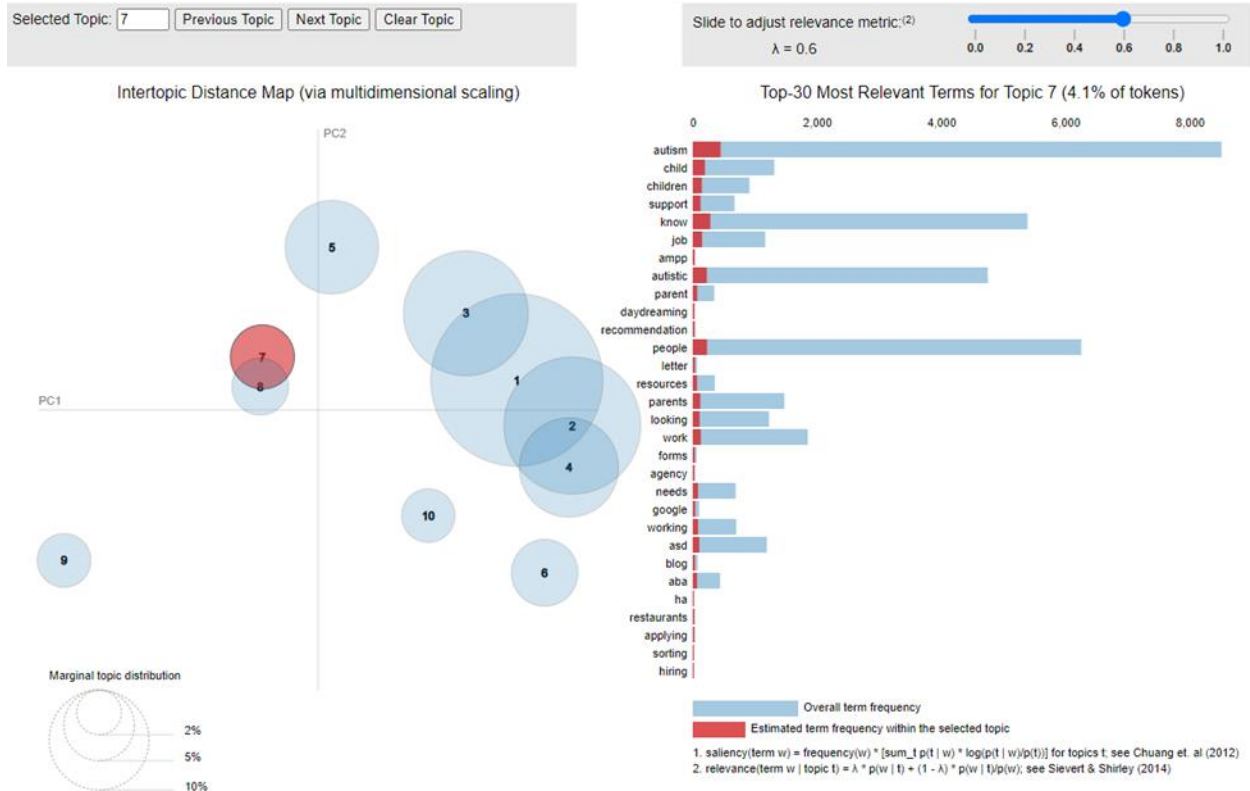
Top-30 Most Salient Terms¹



LDA Most Relevant Topics

Based on the marginal topic distribution of the terms job and work as well as visual analysis of the posts in each topic, the topics most relevant to the question were identified as topics 7, 4, and 6. While topic 1 was also identified, this large and broad topic was not selected as most relevant.

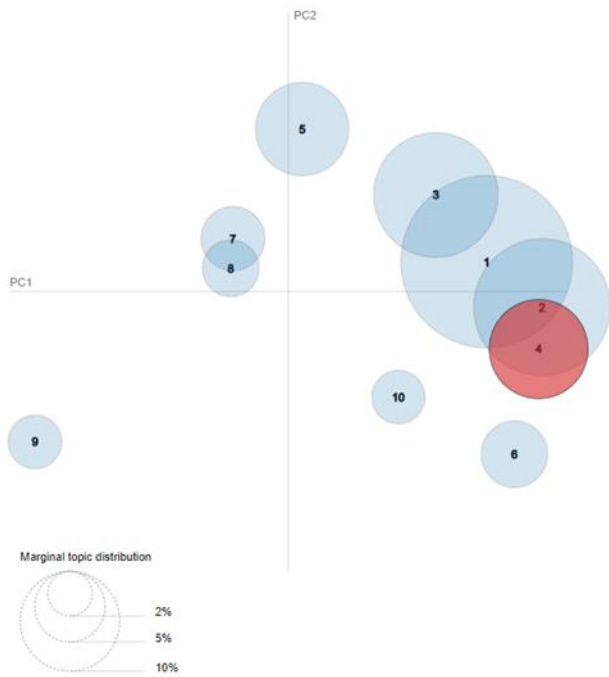
Topic 7 contains several relevant terms for work including job, work, forms, agency, working, restaurants, applying, and hiring. The latter six terms are present exclusively or almost exclusively in topic 7. Topics 4 and 6 only contained the terms job and work as their most relevant terms.



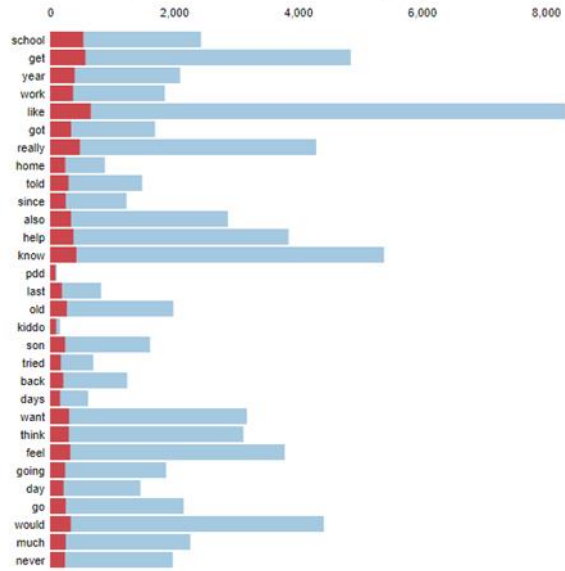
Selected Topic:

Slide to adjust relevance metric:⁽²⁾
 $\lambda = 0.6$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 4 (9.8% of tokens)

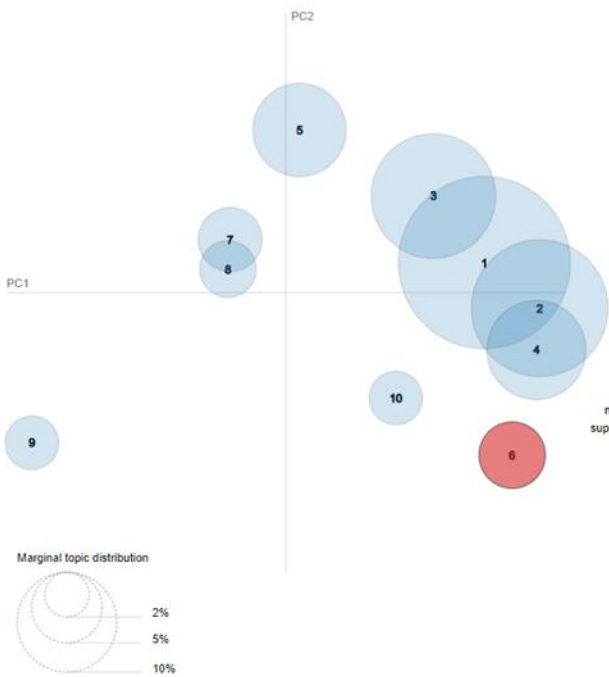


1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
 2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

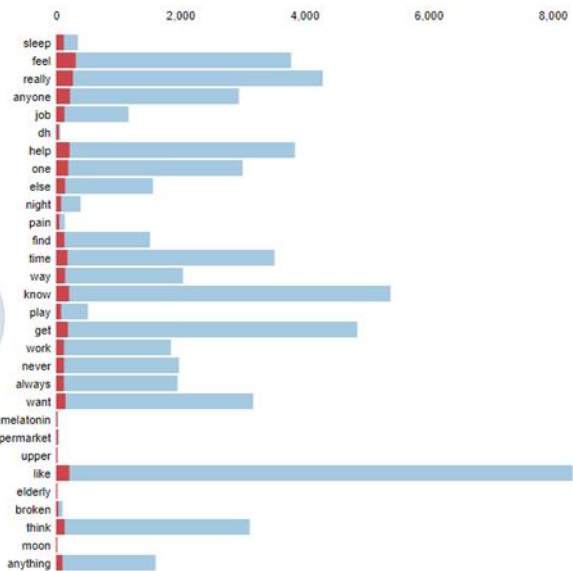
Selected Topic:

Slide to adjust relevance metric:⁽²⁾
 $\lambda = 0.6$

Intertopic Distance Map (via multidimensional scaling)



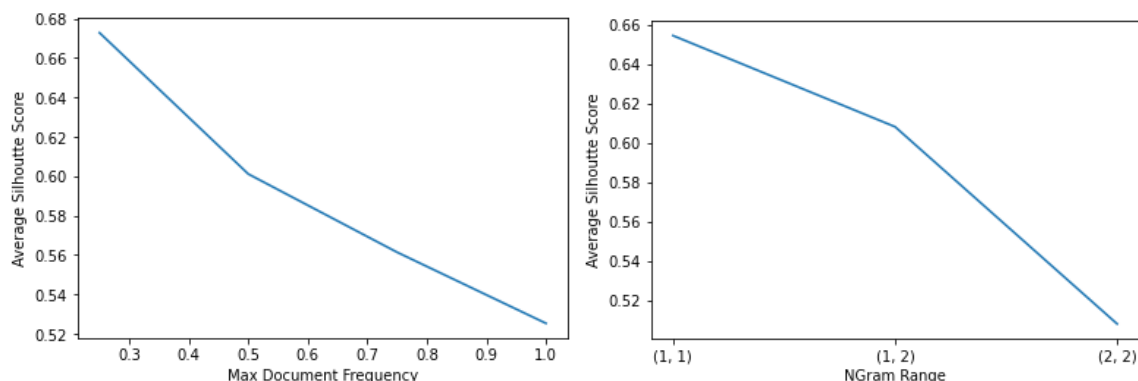
Top-30 Most Relevant Terms for Topic 6 (4.4% of tokens)



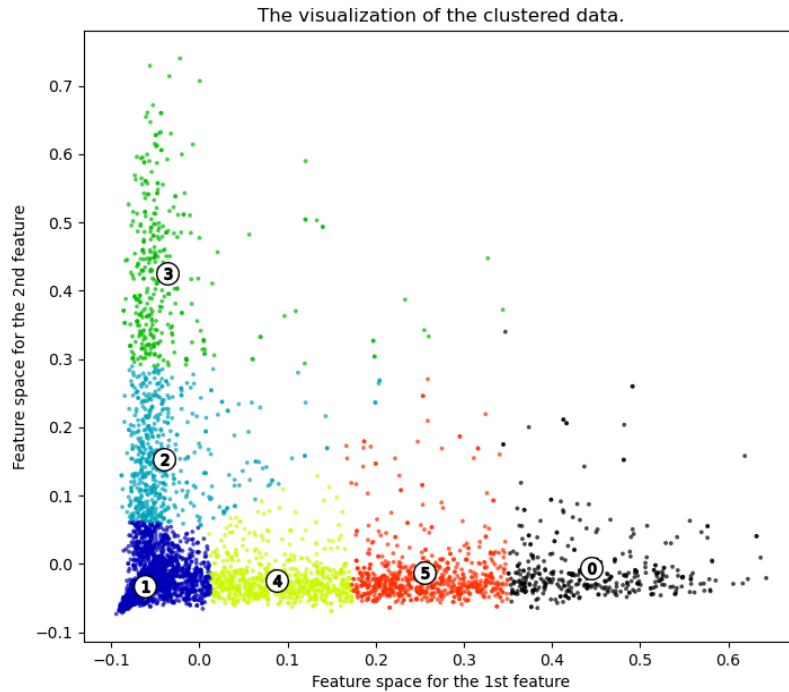
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
 2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)

Tf-idf with K-Means clustering results

Hyperparameter testing revealed that lower values of maximum document frequency and smaller n-gram ranges lead to better clustering performance. This suggests that there are many terms in this corpus which occur frequently across all documents thus excluding a higher number of terms leads to greater differentiation among the documents. This is intuitive given that the documents all come from the same source community, and are thus likely to be very similar.



Given the results of hyperparameter testing, clustering was applied to the tf-idf representation obtained from the model with maximum document frequency of 0.25 and n-gram range (1,1). A qualitative analysis of the top terms and phrases from each of the resulting clusters was performed to determine the optimal number of clusters. This analysis revealed that six clusters yielded the most meaningful partition of the dataset, while still maintaining a high silhouette score of 0.68. Clustering results are shown in the figure below. Size of clusters varied from 566 to 3790 with a mean of 1474.67 and a standard deviation of 1148.97. Cluster 1 is identified as being related to employment. This cluster contains all posts that contain the term 'job'. Additionally, the 3-gram phrase 'full time job' is among the top 20 most frequent 3-grams in this cluster.



LDA and tf-idf

Comparison of Groupings

A comparison of the groupings performed by the LDA and tf-idf models was performed. First, a Cramer V correlation was calculated comparing the categorizations completed in each model. The association between unique categories was 4.45 percent, which is a small association.

		if-idf cluster						Total
		0	1	2	3	4	5	
LDA Topic	1	17	47	5	12	19	110	210
	2	5	12	2	4	6	38	67
	3	6	12	3	2	2	33	58
	4	2	23	8	5	8	41	87
	5	49	148	23	43	57	233	553
	6	30	73	27	25	48	168	371
	7	251	629	174	209	229	957	2449
	8	167	403	108	147	222	861	1908
	9	166	475	149	174	229	876	2069
	10	83	255	67	87	111	473	1076
	Total	776	2077	566	708	931	3790	8848

Next, a Tetrachoric correlation was calculated between relevant vs not relevant groupings by each model. The association between relevant and irrelevant categorizations was found to be 4.3 percent, which is a very weak association. In comparing the two models, only 8.19 percent of posts were in a relevant group in both models, 39.94 percent were in a relevant group in only one model, and 51.86 percent were not relevant in either model. Of the 725 posts determined to be relevant in both models, 629 or 86.8 percent of these posts, were in the most relevant LDA topic, Topic 7.

		if-idf cluster		
		Relevant	Not Relevant	Total
LDA Topic	Relevant	725	2182	2907
	Not Relevant	1352	4589	5941
	Total	2077	6771	8848

n-grams

The top 5 *n*-grams are presented from the most relevant tf-idf cluster and the three relevant LDA topics. In the case of ties, all tied are presented with the exceptions of tf-idf Cluster 1 and LDA Topic 4: 5-grams which had 95 and 48 ties respectively. Full results are provided as an additional document. Phrases relevant to employment were not present in the top 5, but were present in inspecting the full outputs.

	3 grams	4 grams	5 grams
tf-idf Cluster 1 n=2077	<ol style="list-style-type: none"> (('high', 'functioning', 'autism'), 110) (('make', 'eye', 'contact'), 45) (('anyone', 'else', 'feel'), 29) (('high', 'functioning', 'autistic'), 26) (('year', 'old', 'male'), 25) 	<ol style="list-style-type: none"> (('diagnosed', 'high', 'functioning', 'autism'), 11) (('anyone', 'else', 'feel', 'way'), 10) (('consume', 'certain', 'pieces', 'art'), 8) (('anyone', 'else', 'feel', 'like'), 8) (('rigid', 'restricted', 'thinking', 'black'), 6) (('black', 'amp', 'white', 'thinking'), 6) 	<ol style="list-style-type: none"> (('never', 'tried', 'make', 'friends', 'always'), 5)
LDA Topic 7 n= 2461	<ol style="list-style-type: none"> (('a', 'lot', 'of'), 359) (('on', 'the', 'spectrum'), 341) (('i', 'feel', 'like'), 261) (('i', 'want', 'to'), 233) (('i', 'have', 'a'), 211) 	<ol style="list-style-type: none"> (('know', 'what', 'to', 'do'), 86) (('i', 'feel', 'like', 'i'), 62) (('i', "don't", 'want', 'to'), 52) (('i', 'was', 'diagnosed', 'with'), 51) (('i', 'just', 'want', 'to'), 50) 	<ol style="list-style-type: none"> (('don't', 'know', 'what', 'to', 'do'), 40) (('dont', 'know', 'what', 'to', 'do'), 38) (('i', "don't", 'know', 'how', 'to'), 31) (('i', "don't", 'know', 'what', 'to'), 28) (('i', 'dont', 'know', 'what', 'to'), 25)
LDA Topic 4 n=89	<ol style="list-style-type: none"> (('i', "don't", 'know'), 14) (('i', 'feel', 'like'), 13) (('a', 'lot', 'of'), 10) (('when', 'i', 'was'), 10) (('know', 'how', 'to'), 9) 	<ol style="list-style-type: none"> (('i', "don't", 'know', 'if'), 6) (('on', 'the', 'autism', 'spectrum'), 5) (('know', 'what', 'to', 'do'), 5) (('i', "don't", 'know', 'how'), 4) (('don't', 'know', 'how', 'to'), 4) (('i', 'would', 'like', 'to'), 4) 	<ol style="list-style-type: none"> (('i', "don't", 'know', 'how', 'to'), 4) (('i', 'am', "'not', 'autistic', 'enough'"), 3)
LDA Topic 6 n=370	<ol style="list-style-type: none"> (('on', 'the', 'spectrum'), 52) (('a', 'lot', 'of'), 40) (('i', 'have', 'a'), 39) (('i', 'want', 'to'), 35) (('i', 'feel', 'like'), 25) 	<ol style="list-style-type: none"> (('know', 'what', 'to', 'do'), 11) (('and', 'i', 'want', 'to'), 9) (('i', "don't", 'know', 'how'), 8) (('i', 'was', 'wondering', 'if'), 8) (('i', 'would', 'like', 'to'), 8) 	<ol style="list-style-type: none"> (('i', "don't", 'know', 'how', 'to'), 6) (('and', 'i', 'am', 'trying', 'to'), 4) (('to', 'be', 'a', 'better', 'sister'), 4) (('i', 'dont', 'know', 'what', 'to'), 4) (('dont', 'know', 'what', 'to', 'do'), 4) (('wouldn't', 'make', 'fun', 'of', 'someone'), 4) (('make', 'fun', 'of', 'someone', 'who'), 4) (('fun', 'of', 'someone', 'who', 'has'), 4)

Containing Keywords

Each post was tested for containing partial keywords, job and work, the two most identified terms from LDAvis related to employment. The presence of each term was binary coded separately and summed for the post.

Average presence of the two terms is presented. Red represents the bottom quartile, yellow represents the mean, and green represents the top quartile of term presence. tf-idf Cluster 1 scored the highest with an average presence of 0.50 per post. LDA performed poorly as maximum averaged total presence was 0.33 terms per post.

		tf-idf Cluster						Total
		0	1	2	3	4	5	
LDA Topic	1	0.18	0.68	0.40	0.33	0.21	0.15	0.30
	2	0.20	0.33	0.50	0.25	0.17	0.05	0.15
	3	0.00	0.42	0.33	0.50	0.00	0.09	0.17
	4	0.00	0.43	0.63	0.20	0.50	0.24	0.34
	5	0.33	0.41	0.35	0.42	0.23	0.15	0.27
	6	0.17	0.58	0.59	0.28	0.27	0.15	0.29
	7	0.25	0.48	0.36	0.44	0.29	0.22	0.33
	8	0.23	0.49	0.45	0.41	0.35	0.21	0.32
	9	0.20	0.51	0.34	0.46	0.38	0.21	0.33
	10	0.25	0.54	0.24	0.45	0.29	0.15	0.30
	Total	0.24	0.50	0.37	0.43	0.32	0.19	0.31

Discussion

The ten-topic LDA model identified one highly relevant topic and two other relevant topics, totalling 2907 posts most likely relevant to employment, while the six-cluster tf-idf model identified one highly relevant cluster, with 2077 total posts most likely to be relevant to employment. The intersection of these document subsets yield 725 posts or 8.19 percent of all posts.

Comparing the presence of most relevant terms identified from LDAvis, work and job, suggests that tf-idf was better at topic discovery related to employment for this dataset as the three most relevant subsets, tf-idf topics 1, 3, and 2 respectively, all came from the tf-idf model. Surprisingly, the three grouping combinations which produced the highest overall relevance (tf-idf cluster 1/LDA topic 1, tf-idf cluster 2/LDA topic 4, and tf-idf cluster 2/LDA topic 6 respectively) were not from any of the overlapping relevant groupings. This suggests that the combination of these models yield more meaningful partitions than either of the models individually. Thus, use of both approaches is warranted for any future work.

While this project found promising results there are some limitations related to the attributes of the dataset. First, the corpus size is relatively small at only 8869 total documents. This problem is likely a result of the data collection which gathered only parent posts from reddit and did not gather comments that responded to the parent posts. Future research should gather both parent posts and comments to increase the sample size of the corpus and capture a more holistic view of discussion in the community. Finally, there is no existing research that we are aware of regarding the demographics of users on r/autism or the proportion of all employed or job seeking adults with autism who use reddit and/or post to r/autism. Therefore, it is impossible to say how representative this dataset is of the autistic population at large.

The present project utilized topic modeling and document clustering to partition a text corpus from r/autism and conducted a mixed methods analysis of the resulting document subsets. We compare the results of two approaches, topic modeling with LDA and document clustering with K-Means clustering applied to a tf-idf document representation. Findings show that tf-idf was superior to LDA, but suggest that there may be underlying interactions between the groupings of additional posts of interest which presently elude each model separately. When groupings of each model are applied, the larger dataset is able to successfully be partitioned for further analysis of employment. A hand review of these overlapping posts shows promising results for the validity of these topic discovery methods. Further analysis of these posts can provide researchers and practitioners with unique insights into the experiences, challenges and barriers surrounding employment for adults with ASD. Future work in this area can expand this corpus by gathering additional reddit posts, as well as comments posted in reply to those

posts. Additionally, these techniques can be implemented in an online fashion in order to track changing trends in discussion among this community. Future work in this area can expand this corpus by gathering additional reddit posts, as well as comments posted in reply to those posts. Additionally, these techniques can be implemented in an online fashion in order to track changing trends in discussion among this community.

References

- [1] T. Barkin, "Breaking down the labor shortage," Federal Reserve Bank of Richmond, Richmond, Virginia, March 3, 2022.
- [2] C. W. Munk, "Tech companies say they can't find good employees. The companies may be the problem," in *The Wall Street Journal*. [Online]. Available: <https://www.wsj.com/articles/tech-companies-say-they-cant-find-good-employees-the-companies-may-be-the-problem-11615212029>
- [3] R. D. Austin, and G. P. Pisano, "Neurodiversity as a competitive advantage," in *Harvard Business Review*. [Online]. Available: <https://hbr.org/2017/05/neurodiversity-as-a-competitive-advantage>
- [4] C. Solomon, "Autism and employment: Implications for employers and adults with ASD," *Journal of Autism and Developmental Disorders* 50 (2020): 4209-4217.
- [5] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 5th ed. (2013). <https://doi.org/10.1176/appi.books.9780890425596>
- [6] C. Sievert, and K. E. Shirley, "LDAvis: A method for visualizing and interpreting topics." *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces* (2014): 63-70.
- [7] Shah, Neepa, and Sunita Mahajan. "Document clustering: a detailed review." *International Journal of Applied Information Systems* 4.5 (2012): 30-38.
- [8] Koopman, Cynthia, and Adalbert Wilhelm. "The effect of preprocessing on short document clustering." *Arch. Data Sci. Ser. A* 6.1 (2020): 01.
- [9] Curiskis, Stephan A., et al. "An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit." *Information Processing & Management* 57.2 (2020): 102034.
- [10] Barker, Joshua O., and Jacob A. Rohde. "Topic clustering of e-cigarette submissions among Reddit communities: A network perspective." *Health Education & Behavior* 46.2_suppl (2019): 59S-68S.
- [11] Bellon-Harn, Monica L., Ryan L. Boyd, and Vinaya Manchaiah. "Applied Behavior Analysis as Treatment for Autism Spectrum Disorders: Topic Modelling and Linguistic Analysis of Reddit Posts." *Frontiers in Rehabilitation Sciences* (2021): 102.
- [12] Dansby, Rachael A., et al. "A phenomenological content analysis of online support seeking by siblings of people with autism." *Journal of Family Psychotherapy* 29.3 (2018): 181-200.
- [13] D. M. Low, L. Rumker, J. Torous, G. Cecchi, S. S. Ghosh, & T. Talkar, "Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on Reddit during COVID-19: Observational study." *Journal of Medical Internet Research*, 22.10 (2020): e22635.

Table 1 Randomly Selected Posts from Top Groupings and Grouping Combinations

Topic	Sample Posts
LDA 4.1	<p>struggling in general hi im 23/f from the uk im currently pursuing a diagnosis im still waiting for a psych referral it can take up to 18 weeks ive known for all my life that i can remember of course that i was different but it wasnt until the last couple of years after having got to know a few people with asd that i considered speaking to my gp about it ive done lots of research spoken to lots of my friends who have asd spoken in forums etc and i finally feel like i may have an answer which is great because i can learn to better cope from day-to-day with the correct support i could receive a different diagnosis who knows but im pretty sure i wont ive really struggled in the past few months ive had some kinda traumatic events and i feel like theyve pushed back my mental state as a whole im depressed lethargic demotivated most of the time ive been drinking too much in order to have the confidence to go out and socialise rather than hiding away in my house that then progressed to drinking while hiding away in the house so i can get some sleep and not feel what i dont want to feel i had a meltdown at work a couple days ago its been brewing for a while and ive been exhausted from holding back for a long time ive been open with my boss in terms of my doctors appointments etc and whats happening so i had hoped she may understand me a little better so i had this meltdown and was actively trying to remove myself from the situation but she kept following me with every step away i took and was shouting in my face and pointing in my face i kept asking her to stop there were customers in the store also who had turned around in their seats to watch i took myself into a quiet cooler room to calm down and she kept coming at me telling me to get out but i was crying so much i could barely see through my tears and my face was covered in mucus yum i didnt really want to walk into a busy main street like this everything was just too much i really wish i could have exerted better self control over myself at the time but i just couldnt and once it started i couldnt stop it so we spoke the next day im not allowed to work for the next week and i get paid hourly and she says she doesnt understand me and never will i get it shes not a psychologist i dont really know what the point in this post is if im writing to get it off my chest with people who will maybe understand or if im asking for advice obviously i dont have an official diagnosis so i think legally she doesnt have to make any adjustments etc for me but i feel like ive been as open as possible with her to ensure we can work cohesively and effectively i think shes under a lot of stress too theres a lot happening at work and i dont think our jobs are safe she can be very sharp and sometimes very intimidating she also uses blackmail and she always has her favourite and then another flavour of the month i do like her in general but sometimes she makes it really hard for me to she probably feels the same about some of us idk is there anything i could say that might help her understand more i already explained some stuff on my mind but not the major stuff i dont want her thinking im trying to pull the sympathy card i just know itll open a whole can of worms in my personal life and i cant put myself through it i just feel so worn out and useless recently and i feel like i cant do anything right for anyone sorry to be so doom and gloom</p>
LDA 4.2	<p>in a relationship with an autistic person need some advice i know that autism is a little different for everyone but i figured to ask the autistic community for advice my so is autistic and lately has been having a lot of trouble he's been waking up over stimulated and having more meltdowns non-verbal violent and overall not present does anyone have any advice as to how i can be better for him what would you guys like personally for a loved one to do if you were experiencing something like this thanks for the help sending love to all</p>
LDA 4.3	<p>new here hi everyone i am new here i am 24 i am not formal diagnosed but i was told by my special edu teacher that i may have asperger's some of my special interests are high end audio music and photography i prefer my routine schedules i get a little bit irritated if i don't know if there is going to be a change in my schedule also sometimes i feel like items more than people i also fear talking to people because planning of what to say and stuff anyways i hope to making some friends around here</p>
LDA 4.4	<p>three year old son autism diagnosis my three year old son went to an observational screening this past week is almost certainly going to be diagnosed with mild to moderate autism we have our rural very small school putting together a group of enthusiastic but let's say probably a little inexperienced</p>

	<p>professionals to get a game plan going on how best to proceed at a meeting tomorrow currently he's working with the school under a developmental delay diagnosis due to his lack of speech he jabbars all the time says singular words often but only very occasionally does he say two together i am going to look into getting him into additional speech therapy school district only sees him a couple times a month for about an hour honestly if it weren't for his speech we'd probably never have even considered this possibility i can go into further details on his behaviors etc if it's helpful but in short guess my question is what else should i be doing right now what questions do i need to ask should i be contacting a specialist i just want to make sure i'm doing everything i can to ensure i give my boy what he needs to reach his full potential appreciate the help/support</p>
LDA 4.5	<p>looking for freinds i am looking for freinds who share common interest with me i am interested in video games game developement and computers want someone to whom i can chat with nothing special just feeling lonely</p>
LDA 6.1	<p>adhd + autism anyone struggling with these two commorbid disorders at the same time how did you manage i seriously need tips</p>
LDA 6.2	<p>i'm afraid i'll never get another job i'm 17 and everyone else my age seems to get jobs so easily they just apply go for an interview and bam meanwhile i apply and get no reply and even when i have had interviews i don't know how to answer simple questions and screw it up i held a job for 3 years before this that i got through a family member and i was told i was a good employee but i can't show that to employers now my inability to look them in the eye talk about myself etc i want to go to university to be a teacher of english but i feel that i'll fail at that too how will i teach kids to read and communicate well if i myself can't communicate successfully</p>
LDA 6.3	<p>website recommendation for evidence based facts/medicine regarding autism looking for an easy to read and navigate website to pass onto a friend with an autistic child which reviews up to date research about autism its hard to find good websites among all of the pseudo science ones</p>
LDA 6.4	<p>nt tag can anyone fill me in on how i got a neurotypical tag next to my username in this subreddit</p>
LDA 6.5	<p>4 year old minimally verbal my 4yr old minimally verbal son just said daddy for the first time made me cry sorry just a proud dad who had to share</p>
LDA 7.1	<p>share your stories about loneliness hello everyone! help run a support group for and by autistic people and we have meetings every other week where we discuss a subject linked to autism we also do fun things tomorrow i get to lead a discussion about loneliness i've done my research but now i'm looking for stories so if you have stories about being lonely maybe with advice on how to overcome this i would love to hear them</p>
LDA 7.2	<p>on this new years eve autistic me fucked around with a bell tower that was made in the year 1289 in holland and it was fucking amazing please excuse my english i am dutch and yeah with autistic me i mean i really have autism also the part where i fuck it up begins at the end of the story so kind of a small thing that happened but i got my ass beaten by someone that worked therein the netherlands theres this tradition in the province groningen to make the bell tower ring from 1700 untill 0800 from 2018 to 2019 non-stop without breaks you can visit it and its a very small room theres 2 ropes you can pull on it has to be seperately pulled or else youll fuck up the entire tower and itll crush everyone thats beneath itso theres this very typical guy pulling one rope that one cliché old guy with his white mustash and white beard and blue overall hes obviously been doing this for years one hand in his pocket one hand pulling the rope and you could pull the ropes yourself if you wanted to so my brother-in-law that was with me pushed me and said go ahead its alright so i stood there holding that rope i was holding it for just 1 second i had to act fast or i fucked up the entire tower so i gave it the biggest yank i could ever give but heres the complicated part i got scared and i let go and i grabbed it again but from higher up so what do you think that happened with mei flew up 10 feet in the air while holding the rope someone that worked there the room was seriously very small everyone looked at me</p>

	<p>right away yanked me right from the rope as i came down again and he took atleast 10-20 seconds to synchronize the bell tower again he gave the rope to someone else that would pull it and he gave me the angriest look you could imagine get the fuck out of this bell tower get out now no now he had a few drinks down his throat but i do have to admit all those people who were mid-drunk were doing a goddamng good job at maintaining peace at the place so i went out and my brother-in-law is a very serious and diciplined person but he came out with the biggest laugh i have ever seen he said and i qoutesebastian je bent mijn verdomde held=sebastian youre my goddamn fucking hero and shook my hand it wasnt worth it not in the slightest luckily i live 70km away from that religious place but it was the best time ive ever had in year despite the tower that could have been broken down cuz of me and just to clarify my brother-in-law was there to protect me cuz of my autism as you can see im really impulsive but he didnt drink anything and drove me sober save back to home</p>
LDA 7.3	<p>what do you think my little brothers future might look like hes 4 so some things he can do *** count to 50 a couple months ago when everyone thought he could only count to 20 i heard him by himself counting to thirty and even doing some 30s so i guess he may have learned this by himself *** recognize some colors and shapes *** read some stuff i dont know the extent of this but earlier this year my dad had a suspicion he was able to read low and behold i randomly gave him something at the supermarket right before turning 4 and he read the name to me after a few months we know for sure he definitley can read some stuff *** may be able to spell out his name not sure but my dad said he could *** say the alphabet *** answer some questions that dont require deep thought for example my dad can show him a pic of someone in our family and be like whos that and hell answer things he has trouble with *** paying attention he cant seem to pay attention for more than 5-10 secs and doesnt always respond to calling his name even if you do it loud *** verbalizing what he wants hes getting better for sure but he wont be like i want milk though yesterday he wanted vegetable chips and walked around repeating vegetable chips for a few minutes *** interacting w people *** some noises like the blender makes him cry / scream and run away *** talking he cant have a conversation cuz he cant seem to say too many words mx he sings to himself and babbles a lot like he says wa wa wa a lot idk if hes trying to talk and cant and interestingly enough today my dad asked him if he was ready to leave and he said yes so thats pretty cool any idea what his future may look like do you think he may be able to talk somewhat normally and hold a conversation</p>
LDA 7.4	<p>thread autism and substance abuse my story i was raised in a strict christian family so i wasn't offered as much as a beer before turning 18 when i turned 18 i got a car and commuted to the local community college i'm a guy and was friends with a few girls but never had any guy friends as a kid a large part due to being shy and awkward but having a lisp and being into poetry and indie rock at a fundamentalist christian school didn't help i met a guy in my college class i clicked with and he invited me to his apartment to hang with his roommates they offered me beer cigarettes and a bong rip my first time there within a month i tried coke mdma adderall and shrooms luckily i never got hooked on any of the harder stuff mostly because i refused to possess illegal drugs and only used them when others had them however i got a medical weed card and high dose adderall prescription when i turned 21 and added unlimited alcohol access to the mix i became a complete mess i flunked out of school lost my job and my few friends i got sober about a year ago and have been slowly building my life back up i was diagnosed autistic about the same time sadly my first positive human interactions involved substance abuse and i'm still learning how to connect with people on a sober level i would love if others on the spectrum could share their story of substance use if moderate alcohol or weed use helps in social situations don't be afraid to share</p>
LDA 7.5	<p>anybody else have trouble recognizing an unreliable narrator i never seem to be able to recognize an unreliable narrator in a story until its explicitly pointed out to me sometimes i might see it if theyre making a factual error that i recognize but otherwise im pretty clueless i suspect this goes along with having to accept peoples words at face value i like fiction but only fairly straightforward stories no unreliable narrators no ambiguity about what happened anybody else like this out there</p>
tf-idf 1.1	<p>never had a severe nonverbal episode before and it was kind of scary i've had little bouts of being nonverbal where i'll just be sitting with my boyfriend and hum or shrug in response to his questions</p>

	<p>but never to a point where he might think something was wrong today was different i've had a recent trauma that i'm still recovering from and we've had my sister's bird over while she's on vacation for the past 11 days he is constantly screaming it really just made everything so awful it felt like everything was so loud and i was having a panic attack and then u just couldn't talk i was lying on the couch meditating bc its the only thing that helps my anxiety and the bird was screaming so i couldn't meditate my anxiety away and it just all piled up and i start sobbing but i cant even tell my boyfriend what's wrong so i have to skype message him and then ask him to explain it to my momi think my parents have absolutely no idea how fucking awful sensory overload is 11 days of the loudest fucking bird in the world it's been horribleidk what exactly i'm looking for maybe advice but mostly support or something</p>
tf-idf 1.2	<p>double-speak it's hard for people to understand my words because i often can't put my thoughts into a coherent sentence so people will often misunderstand me and if i'm in a debate the opposing person will use my vagueness against me to try to make me sound like an idiot even if they have nothing to gain from being right or wrong they literally twist it to fit their ignorant arrogant ego anyone else have this issue it angers me because i can understand what i am writing/saying just fine but other people completely skip over it and just go to a default response like i'm speaking spanishedit edited this post 2 times and counting for better understanding</p>
tf-idf 1.3	<p>are people able to figure out that youre autistic there are some people in the world who have the ability to tell whether someone is autistic by the way the person walks/talks so have you ever met someone who figured out you were autistic even before you told them</p>
tf-idf 1.4	<p>feeling left behind i know i'm not the only person who feels like this but because of the way i am i tend to be more childish than my younger brother and such which can painfully show when he and my mother are talking about bills and general more down to earth topics of the likethis bugs me for everytime they talk about something like this i just happened to want to talk about trivial things like going outside or playing a video game together something more light hearted but i tend to be overshadowed in these kinds of situations i feel and it just ends up having me feel depressed in the end my brother is younger than me yet i feel he's "out growing" me so to speak and it's really taking its toll on my mental state i feel like i'm going to be a kid forever and i'm really starting to hate it more and more it happens the feeling "out grown" and childish in normal situations i just feel like i'm going to be a kid forever and i really do hate it i can't entirely just stop being the way i am but to be frank i understand how a certain character feels like more and moremy mother doesn't entirely treat me like my age 23 for i guess it's because i live with her and my brother but still it just doesn't really feel right i wish i could describe it more but tight now i'm too depressed to think in depth</p>
tf-idf 1.5	<p>humanities courses in college i'm a first-year mechanical engineering student diagnosed with asperger's earlier than i can remember the university i attend has a mandatory number of humanities credits in order to graduate which i've decided to cover with an honors sequence and despite the fact that i got good grades in last semester's course i always feel like i struggle with it even into the second semester it causes me a lot of stress to the point where i've had minor meltdowns over certain major assignments i don't have anywhere near this level of difficulty with my science and math courses so i'm starting to wonder if it's just an autism thing did anyone else have similar difficulties if so how did you deal with them</p>
LDA 7 and tf-idf 1 (top group from each) 1	<p>paralyzed long post need advice i'm 16 and i'm going to be turning 17 in a couple of months i've recently been experiencing this feeling that i'm losing control over my life and everything is so overwhelming it's like everyone else my age is so far ahead of me and i feel like i'm still a kid in some aspectslet me explain i've been struggling with a lot of things recently i've had depression and anxiety since i was 9 or 10 and i've had trauma since i was 4 or 5 my dad has tbi traumatic brain injury and my mom died last year my parents divorced when i was 5ever since i was little i've felt this need to take care of others and my dad however i have an inability to take care of myself i have so many things i have to do - get a driver's license figure out which colleges i want to go to get a job sign up for the act take care of my ap classes etc i really want to take care of the first 3 things but i</p>

	<p>feel like i'm literally just stuck for one i'm terrified of driving i'm really bad at processing everything on the road at once and i've almost bumped into several people i just don't trust the drivers on the road because i don't know what they're going to do and i don't trust myself because of my sensory issues i want to figure out what college i want to go to so i can feel more prepared for the future but i also feel terrified of this too because i feel trapped i keep thinking "what if i end up broke" "what if college is useless" "what if i'm pursuing the wrong thing" stuff like that my dad says i'll be getting grants and scholarships but i doubt it will cover all of my college expenses my biggest thing is struggling to push myself to get a job i had a job for a short time but my experience was terrible i have hearing loss and combining that with sensory overload working in a large fast food restaurant where the managers are socially political was not a good experience i want to work for the sake of working and nothing else i barely care about money and maybe that's a problem i don't know and it seems i am incapable of pushing myself to get a job because nearly all the jobs that i want and would be the least overloading are not available and jobs that are available i just don't care about or they're overloading i don't get why everything has to be like this it's completely beyond my understanding i just want to be a psychologist when i grow older but there's so many obstacles in the way whenever i accomplish something it takes weeks or even months for me to recover because everything is so tiring i've even thought about what it might be like to just end my life here i already feel like i'm behind how could i possibly deal with things later in life i need advice i know i need to get moving on things but i just feel paralyzed i don't know what to do it's like i've lost my purpose any advice would help at this point tldr i'm 16 and i feel like i'm stuck like i can't get myself to move and keep the momentum going when i do accomplish something i lose energy and take weeks to recover any advice would help</p>
<p>LDA 7 and tf-idf 1 (top group from each) 2</p>	<p>so i went to the global climate strike im home right now my throat sore feet hurting from walking that much and really really tired first i didnt really want to go there- i hate to feel sweaty with ppl all around me screaming i think a lot of you can relate but it was surprisingly great and fun and i care about the earth a lot the people there were nice as well- we talked to a group of boys from the british school in our city i live in tokyo even though i didnt talk to them only my friends did i felt and still feel pretty accomplished- not proud though i wish i could talk to strangers</p>
<p>LDA 7 and tf-idf 1 (top group from each) 3</p>	<p>can someone explain how neurotypical view the world hii've only recently a week ago learned that i am on the spectrum which explains a lot of things i've struggled with my 40 years of life one thing i've always struggled with is trying to figure out how others think i've always known i think differently then most now i know why and i can find tons of people explaining how someone that is on the spectrum tend to view things but while that's nice in seeing some recognition what i'd really like to have a more clear view on is how the "rest of the world" views thingsbecause it's really a struggle to live in this strange world for me people just don't make sense to me often they seem stupid ignorant foolish emotional rude untrustworthy and so forth and i'm sure they're not or at least not as much as they seem to me i'm just missing some dimension in which others tend to think or work at least that's my guesso yeah not sure if someone has ever found a post or such that goes into this or if someone can explain it even a little i think it would help me</p>
<p>LDA 7 and tf-idf 1 (top group from each) 4</p>	<p>i am very isolated sorry for breaking any rules for this subreddit so i am studying physics and i live pretty far away from my family for almost 3 years nowi am not diagnosed with anything and i think in uni i am doing fine though the test phase will be hard but in the last 3 years which i think is a long time i basically just completely isolated myself and usually i act quite strange obsessing over noises and light and so on and so oni just think there is something very wrong with mebut i already made an appointment with a psychologist from my university i am writing this because this semester i am very suicidal and i don't like talking to my family</p>
<p>LDA 7 and tf-idf 1 (top group from each) 5</p>	<p>getting upset over dumb things i know that if i were any other person i wouldn't get so upset over such trivial things but i can't help it how do i work with thisfor example my dad and i were going to binge watch two seasons of agents of shield within about forty-eight hours because that's all the time he had off work it's my absolute *favorite* show i was so excited and we got through season 4 in less than twenty hours and we were moving quickly through season 5 but then he decided to sleep in and</p>

	roll around in bed and i've done the calculations like eight times and i know *for a fact* that we won't have enough time to finish season 5 to the finale before it's time for me to go back to my mom's place i'm *so upset* so upset that it's giving me heart palpitations and making me feel physically ill and i'm noticing the beginnings of sensory overload because the lights are bright and every sound is irking me and i'm just *so frustrated* because i really *really* wanted to finish the season today with him and now i can't he'll watch it without me but the thing is is that it's something trivial it's not that important but i can't help but be *so damn mad and angry and sad* and this is just one of *many* things sometimes i'm able to control it sometimes i'm not sometimes i'm overworked and i need to just *leave* and get out and sometimes i'm able to pull it together how do i get around this how do i think rationally how do i work through these irrational frustrations
LDA 1 and tf-idf 1 (highest total keywords) 1	my friend has high functioning autism can someone contribute their perspective if this isn't allowed here someone tell me and i will remove it so a friend of mine has autism i've done some research on autism so i have an okay understanding of it however i really really want to understand this from the perspective of people who have it as well because i don't have it i will never understand it but i want to do everything i can to help out and support him what are some things i should know are the things i shouldn't say or should say what are the best ways i can support him in general literally anything would be appreciated stories opinions advice perspective etc thank you so muchh
LDA 1 and tf-idf 1 (highest total keywords) 2	how much sleep do you folks get i'm a 26 year old male who was diagnosed a couple of years ago i understand that it's common for our energy levels and sleep needs to work differently from nts my nt peers don't understand this and it has been making me question how many of us with asperger's and autism require more sleep than nts i typically thrive with about 85-9 hours of sleep if things are busy with school and work i do better with closer to 10 hours of sleep i get criticized for this even by my 28 year old best friend who is used to getting 6 hours of sleep anyway how much sleep do you all feel that you need does it vary by how much you have to do during the day if so by how much thanks everyone looking forward to your replies
LDA 1 and tf-idf 1 (highest total keywords) 3	my social life is completely dead its really bad my family japanese father korean mother moved around a lot and i grew up going to international schools then we moved to japan when i was around 12/13 im 17 now and i never adjusted properly japanese society is generally on the introverted side so when we first moved here i was actually super happy that my classmates left me alone but now im in a lot of f*cking trouble because ive ended up nearly illiterate and i cant hold a simple conversation without choking up or having no idea what to say at this rate i doubt i can even hold a minimum wage job when i graduate im fairly certain this all stems from my autism ive never been diagnosed but at this point its extremely obvious that i have some sort of mental stuff going on and it doesnt help that east asian countries dont give a sh*t about mental health i honestly dont feel comfortable sharing my personal problems on reddit but im kind of desperate id appreciate any advice on what the hell im supposed to do in this situation i feel like i shouldve made a throwaway for this but whatever
LDA 1 and tf-idf 1 (highest total keywords) 4	does anyone else feel ready to cut ties quit a job or walk out of somewhere after a negative social interaction probably my most frequent example i work with some people who tend to get mean for no reason itll go from jokes to purposely striking nerves pointing out im too much of an autist to think a show is funny or saying i blurt things out too much etc and not just saying it once but just repeating it over and over and they warn the new guys in front of me that im a mess or a joke instead of sticking up for myself and then moving on i cant handle it i feel like im gonna have a meltdown as soon as it starts and i shut down i want to quit my job and walk out once i ran to the bathroom and started hitting my head over and over cause i got so overwhelmed ive never been one who cant take a joke or criticism but for some reason when people get a little too much i just cant handle it does anyone else feel overwhelmed like this during something minor
LDA 1 and tf-idf 1 (highest total)	what have been your experiences with cbt and mindfulness i've been very avoidant of any cbt i was supposed to get it with my last therapist but i had to leave therapy with them i'm not sure if i want to pursue cbt with my next therapist due to recent experiences i've been going to a recovery college where they use a lot of mindfulness its also muddled a bit with other techniques based around

keywords) 5	behaviour modification and positive affirmations i'm finding it just too fluffy for me it's all generally hard to do mindfulness meditation is uncomfortable and keeping diaries of emotions etc isn't useful as i'm already very conscious of my emotions i think many of my struggles are more due to autism and adhd than my depression eg if i struggle to do a task then behavioural modification like rewarding myself won't work as the difficulty is due to executive dysfunction not lack of motivation due to depression am i just being put off by the muddiness of what i've experienced so far or is cbt like this in general so possibly not for me
----------------	---