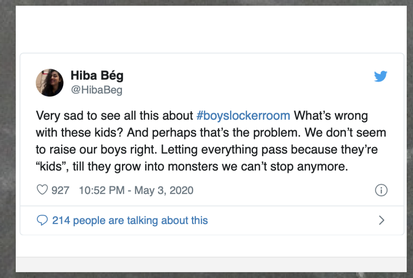
YOUTH PERCEPTION OF SEXUAL EDUCATION IN INDIA

Anusha Subramanian | June 18, 2020| Class: DIGHUM 100 | Instructor: Adam Anderson

While India continues to make phenomenal progress in global fields, there's one particular platform in which it still lacks direction and growth - Education, particularly Sexual Education (Sex Ed). While countries like USA have State specific laws that target the implementation of Sex Ed in schools, until very recently, India was still debating the merits of including it in the curriculum at all. A recent Instagram scandal in May 2020, termed the 'Bois Locker Room' made headlines when a group of adolescent boys were outed by their classmates for a group chat that perpetuated rape culture, objectification of women and criminal behaviours of morphing private photographs of women. While there are many nuances to this issue, it brought to the forefront the need to destigmatize sexual culture amongst teenagers and facilitate dialogue within the education systems on topics that are still considered taboo in the 21st century such as safe sex, gender identity, attraction, consent etc. from more than a paltry biological point of view. In a study conducted by the Indian Ministry of Women and Child development in conjunction with UNICEF in 2007 showed that 53% of children in India faced sexual abuse of some kind and a majority of those went unreported. Recent events and lack of educational reforms implemented lead me to believe that not much has changed since then in the system itself. Research shows that Sex Ed plays a major role in sexual violence prevention. Drawing from that, it is also my opinion that destigmatization and normalisation of sexual topics will lead to better adjusted adolescents who understand concepts of consent & safe sex and will eventually lead to lower sexual harassment cases or crimes.



The responses from an anonymous survey that I conducted after the Bois Locker Room scandal will be my dataset for this project. Target audience for it were people in the age group of 16-28 - individuals who have graduated from the Indian primary educational system (given their 10th grade country-wide board examinations) but would still remember the experience. Sample is restricted to those who had studied in Indian schools (barring distinction on curricula like ICSE, IGCSE etc.) until 10th grade at least.

QUESTIONS

DATA

- DEMOGRAPHIC
  - Age & High School City of Participants → Bar Graph
  - Gender Identity of Participants → Pie Chart
  - Religious Affiliations of Participants → Ring Pie Chart
- PERCEPTUAL/ CURRICULAR (about the curriculum)
  - What topic is rated the highest? → Bar Graph
  - How often are they topics addressed in school? → Heatmap
  - Have participants experienced sexual policing in school? → Bar Graph
- FREE RESPONSE
  - How do you think these topics were approached in your school environment?
    - LDA → Latent Dirichlet allocation to extract relevent topics (TfidfVectorizer())
    - WORD2VEC → Find contextual similarity between groups of words

MATPLOTLIB & SEABORN

CENTRAL RESEARCH QUESTION

What do the youth perceive sexual education in India as?

FRAME OF REF

Organisations that focus on curriculum curation, women's studies and offshoots of the educational board of India are my audience. The frame of reference for this project is childhood within the Indian education system.

EXPECTATIONS

In terms of results, I expect topics generated to indicate the lack of proper Sex Ed coverage in schooling and I also expect the sentiment analysis on the topics to be primarily negative. I think that students with higher religious affinity will rank topics in the survey as less important overall.

SECONDARY SOURCES:

There are no stringent laws in India that link to this so I've been looking at research papers in an Indian context and existing governmental reports in order to gauge the attitude of policy makers towards Sex Ed, its implementations and impacts.

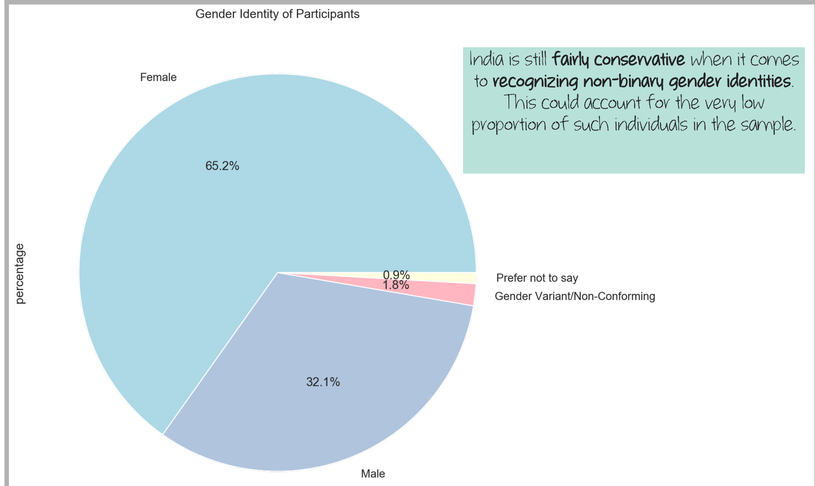
TABLE: SUMMARY OF 'TABOO' TOPICS IN SURVEY

KEYWORD	DESCRIPTION
svsh	Sexual/Domestic Violence and Identifying Workplace Harassment
mental_health	Comprehensive Mental Health Awareness, Assessments & Resources
bullying	Tackling Bullying & Identifying Toxic Environments
sex_ed	Comprehensive Sex Education beyond basic reproduction (safe sex, consent, pleasure, desire)
gender_identity	Gender Identity & Sexuality
menstruation	Normalization of Menstruation
cyber_crime	Comprehensive understanding of Cyber Etiquettes, Crimes, Laws

WORKS CITED

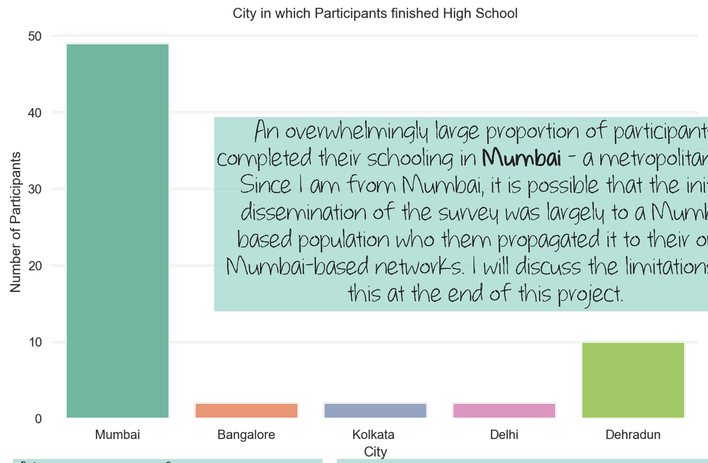
ML resources: Topic modelling with Scikit, LDA visualisation, Gensim Tutorial, DIGHUM100 Topic Modeling Notebook Code, TF-IDF from scratch, BoW vs TF-IDF, Introduction to Sentiment Analysis, spaCy 101
Sexual Education Information: National Policy on Sex Ed, US State Policies on Sex Ed, Why is Sex Ed still taboo in India?, Sex Ed In India: Why, What, When, Where, Whom?, Sex Ed before College, Sex Ed and Sexual Violence

DEMOGRAPHIC SURVEY DATA



India is still **fairly conservative** when it comes to **recognizing non-binary gender identities**. This could account for the very low proportion of such individuals in the sample.

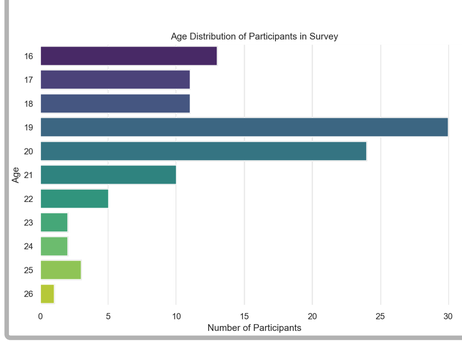
More than half the sample of participants identified as **female**. This could be due to dissemination in gender-skewed networks (I sent them out to more females and they in turn sent it to more-oriented networks and so on), the tendency of particular genders to respond more actively to online surveys than others, or even the cultural need to conform to the binary (individuals who would categorize themselves as non-binary in private were hesitant to do so on a survey).

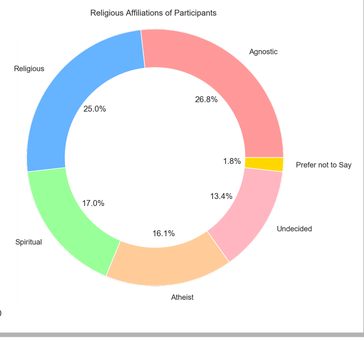


An overwhelmingly large proportion of participants completed their schooling in **Mumbai** - a metropolitan city. Since I am from Mumbai, it is possible that the initial dissemination of the survey was largely to a Mumbai based population who them propagated it to their own Mumbai-based networks. I will discuss the limitations of this at the end of this project.

A large proportion of our participants are in the **19-20** age-group, which means that they very recently graduation from high school. Setting it in context, this was also approximately the age-group that was involved in the Bois Lockeroom Scandal.

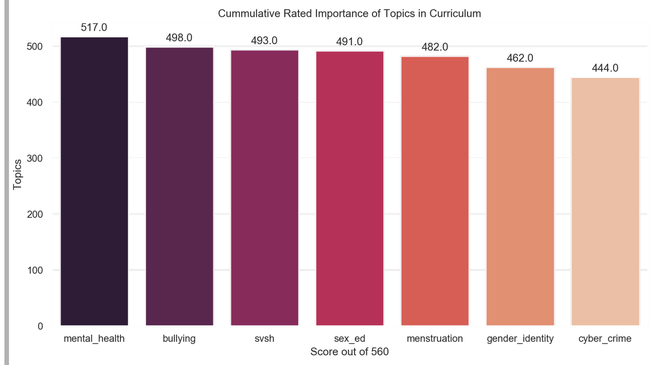
The youth sampled in this survey can be considered representative of the various religious modes of thought prevalent in India. However, I would be careful to assume that they were representative of religions themselves since I didn't ask for particular religious ties such as Hinduism or Islam etc.





PERCEPTUAL SURVEY DATA

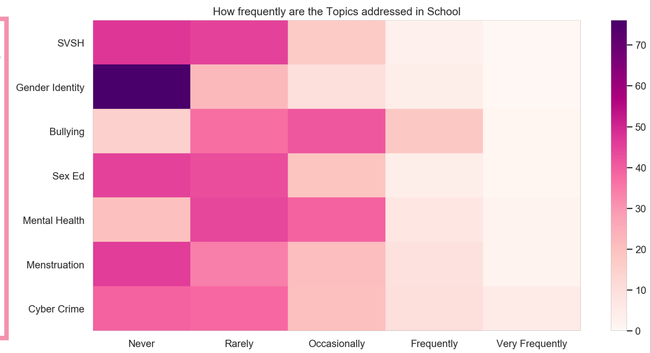
Various forced-choice /rating questions regarding a list of topics relevant directly or indirectly to Sex Ed.



Participants were asked to rate the importance of including the topics listed above in their school curriculum, on a 0-5-point scale. I summed the individual scores obtained by each topic. The graph depicts the cumulative score out of 560.

I expected more variations than this graph depicts. Mental Health and Bullying are the two most important topics according to the participants. However, looking at how close in score the other topics are, it's not surprising because it is possible that a vast majority of teenage/adolescent mental health problems and incidents of bullying are manifestations of issues that arise from improper education and understanding of the topics that rank lower on the graph. Something that struck me as particularly interesting was that Cyber Crime ranked last on the list, even though the survey was conducted in the middle of an episode of national outrage regarding a cyber space scandal.

Participants had to rate on a modified Likert Scale (Never, Rarely, Occasionally, Frequently, Very Frequently) how often the same topics were addressed in their schools. This question along with the previous one was incorporated to understand whether there existed a discrepancy between the desire of youth to gain information on taboo topics and the frequency in which such information was disseminated in the educational institutions.



Topic	Most responses	Least Responses
SVSH	Never	Very Frequently
Gender Identity	Never	Very Frequently
Bullying	Occasionally	Very Frequently
Sex Ed	Never	Very Frequently
Mental Health	Rarely	Very Frequently
Menstruation	Never	Very Frequently
Cyber Crime	Never	Very Frequently

The table above shows that the participants felt that a majority of the topics in this survey were "Never" addressed in their school environment. It is concerning on many levels to hear that basic health conversations such as Menstruation were never addressed, even in an educational sense.

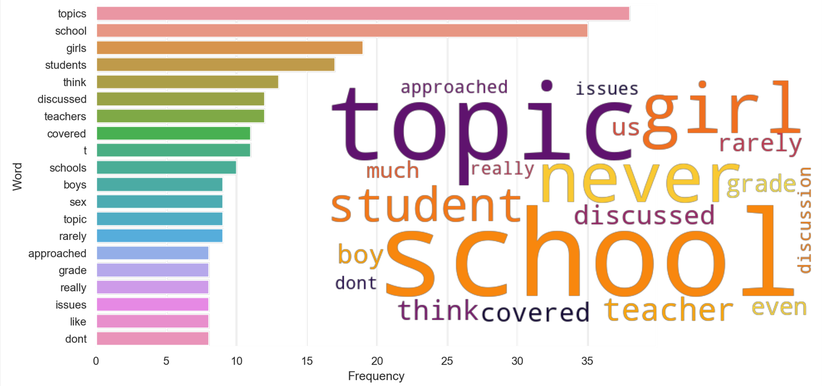
There's almost an even split in the heatmap with the right side dominated by light tones (corresponding to lower number of responses) while the left side is much darker (higher number of responses). Using the given scale to interpret the map, it is evident that very few participants felt that any of these topics were addressed very frequently or frequently.

76 participants felt that Gender Identity was Never addressed in their school - the highest vote in the heatmap. This makes sense given the Indian context. While the Indian landscape accepts the idea of sexual relations, albeit biologically, a large proportion of the country still has trouble with the idea of non-binary Gender identities. And since Education systems and curricula tend to run on a lag, not updating textbooks and details until decades after movement, educational resources about this topic are nearly non-existent in this conservative system.

Since the female identifying participants dominate the dataset, a simple bar plot of responses would not allow for correct interpretation. The extremely high proportion of "Yes" versus the other responses could simply be a function of gender. To get a holistic picture, I plotted the responses in a gender specific manner in a grouped bar plot. **Note**. I only included the binary gender identities since the rest of the categories together comprised <3% of the total participant population.

TEXTUAL ANALYSIS OF FREE RESPONSE

The free responses that I am analyzing in this section are answers to the question of how do schools deal with the topics in this survey. I believe that topics extracted will indicate dissatisfaction and discontent.



One of the primary advantages of a Word Cloud as a visualization is that the frequent words _jump_ out at you almost immediately. It's easy to note some words that take up the most space = 'topic', 'school', 'never', 'girl', 'student', 'discussed' etc. Taken in isolation, they indicate a culture of schools never (or rarely) discussing, 'approaching' or 'covering' such topics as those asked in the question.

We know that participants felt that a vast majority of topics detailed in the survey were 'Never' discussed (Heatmap). This adds additional weight to words such as 'discussed', 'approached', 'issues', 'dont', 'covered', 'think'. This cluster of frequent words, combined with high frequency of 'never' and 'rarely' from this section, and the high rates of 'Never' and low rates of 'Very Frequently' from the heatmap, spin a narrative centered on the characters of 'girls', 'boys' and 'teachers'. If each of those characters independently have a high enough word-frequency to make it on the 'top 20' list, it is reasonable to assume that they were addressed as independent entities in the free responses by participants. I'm hesitant to say that 'girls' and 'boys' being mentioned frequently and separately is indicative of gender disparity when it comes to addressing these topics. Either way, I hypothesis that there continues a narrative of teachers rarely addressing or approaching topics in the survey that students in school are interested in.

WORD2VEC

```
In [44]: model.doesnt_match(["sex","school","topics","discussed"])
Out[44]: 'sex'
```

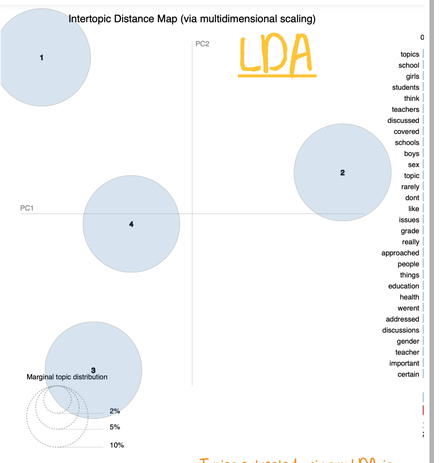
The model picks out word from vocabulary that doesn't fit with the rest of the words - the odd one out. While picking 'sex' as the odd one out from a list of 'sex', 'school', 'topics' and discussed' may not be as simple as saying "that 'sex' was picked as the odd one out because it is not among the topics discussed in school". That would be reductive and exactly one of the main limitations of using this model. However, the fact that 'sex' was singled out as not similar to the rest of the words definitely means something. doesnt_match() returns the index of the document that is furthest away from the mean of all the documents (Word2Vec creates projects the words into vector space and distances). So atleast mathematically, and in this dataset, the word 'sex' is contextually the furthest away from the rest of the words in this document. It definitely is an outlier and while I cannot generalise it to mean that sex isn't addressed in school, I can hypothesize that according to the participants at least, sex is not discussed enough to make them associate it with the other words in the given list, in their responses. And that is part of the overarching point I am making regarding the way youth perceive sexual education in India. What is not included, can be very insightful.

Words that are more than 50% similar to 'sex'

```
In [49]: [word for word in words if model.similarity("sex",word) > 0.5]
Out[49]: ['sex',
          'tried',
          'bad',
          'history',
          'actions',
          'part',
          'sixth',
          'mostly',
          'minute',
          'policing']
```

Note that the word "policing", 'bad', 'actions' and sex have a similarity of over 50%. This could point to the youth perception that sex has been portrayed as something criminal or degenerate in the school and curriculum.

LDA



*Topics extracted using pyLDAvis

#TOPIC	OVERARCHING TOPIC	RELEVANT WORDS
1	Treated as Hostile	crime, obey, authority, harassment, talking, girl, important, topics, school, education
2	Rare Dialogue on Gender & Mental Health	students, covered, rarely, gender, spoken, discuss, tried, mental, talked, importance, different, curriculum, separate, need, planning
3	Taught not Discussed	think, teachers, topic, dont, like, issues, werent, discussions, taught, repercussions, surface, sessions, taboo, didnt, tackle
4	Emphasis on Health	discussed, sex, health, approached, teacher, college, sexual, institution, problems, related, touched, menstruation, mentality, discrimination, opinion, rules, policing

pyLDAvis provided a great way to visualise the topics and their relatedness on a Principal Component Analysis scale (a method for dimension reductionality which still maintains the importance of features by weight). I settled on 4 topics because of my small data set. This was the optimum number which gave me well-distributed topics. Anything higher resulted in severely overlapping topics.

DISCUSSION

LIMITATIONS OF DATA

To talk a little more about the assumptions I have made, it is important to keep in mind that the population sampled in this survey would not be **absolutely representative** of the Indian population OR education system. The demographics of the survey participants show that they are primarily from well-developed metropolitan areas (Mumbai). India is extremely diverse as a nation and there are many pockets even within uniform states where the differences (both ideological and tangible) between the rural and urban are massive. Additionally, the female dominated gender identity in this data set could have affected many parts of the responses such as the ratings of importance of the topics to be included in school curriculum as well as the free response data. I believe females would have naturally voted topics such as Sexual Harassment in Work-place etc. higher than others since it is directly relevant to their gender identity. Similarly, a large proportion of the textual data could have been dominated by female driven experiences.

The biggest problem that I had with **Word2Vec**. Every time I ran my model, it would give me different outputs. A standard analysis become near-impossible. I tried to modify the parameters in order to run the model in a deterministic manner (fixed worker = 1, seed, custom hash function etc.) However, nothing worked. According to FAQs on the RaRe technologies GitHub, Word2Vec does that when the data used to train the model is too little. In this case, I thought it would be best to include some outputs that I had received while working on my project and analyze them.

The Perceptual data indicates a strong trend of the youth perceiving lack of information in essential topics related directly or indirectly to sexual education, particularly in the frequency with which they are covered in school. The topics extracted from the LDA Model serve to show the discontent with the current educational curriculum regarding these topics. It is not possible to ignore the fact that youth aren't getting the sexual education that they want in India, and policymakers can no longer hide under the excuse that the students don't want to learn it. I hope my project serves to highlight that not providing accurate, frequent information to the youth on topics that they believe are important in a school curriculum will lead to them turning to less reliable sources such as the internet and entertainment media in order to fill that information gap. This is detrimental to their development into responsible, openminded and well-adjusted adults.

An important insight for me was the merits of keeping data analysis always at the back of mine while designing surveys or interviews in order to get the best formatted data that will likely save you time when you begin to work on it. When I started this survey, I had not thought through what I hope to achieve by the end of it and thus received data that required manual clean up and standardization before I could begin to process it.

While the survey may not have been completely representative of the target population i.e. the youth of India, it does satisfy its aim of gathering data for a _preliminary_ survey on youth perception. In my opinion, the Demographic and Perceptual Survey Data (the first sections of the notebooks) should be given the most importance due to their objectivity and forced-choice answer design. The textual analysis of the free response should be used as a supplementary resource to back up the results obtained from the first two sections. This would be the best course of action since the Textual Analysis provides us with a "Big Picture" view of the entire dataset along with interpretations of word frequency and similarity and the Demographic/Perceptual Data gives us a metric to weigh the fissure that exists between what the youth want to know and think is important, and their perception of how much of it they are getting in their school environments.

FUTURE DIRECTION

The next step would be to present the project and data to Educational Reform institutions in India or Educational Technology companies to make them aware of this knowledge gap that seems to exist within the youth community in India. I believe that they would have the additional resources and expertise that I could use to carry out more surveys and eventually use the data to curate a State (or Central Government) mandated sexual education curriculum for the Indian Certificate of Secondary Education (ICSE) Board, that is dominated by content that the youth is interested in learning about. This curriculum will require faculty to undergo training and participate in work-shops to sensitize them to issues that they are out of touch with, before they are allowed to teach them in schools.

WORKS CITED (CODE)

Muzzall, Evan. Notebook Week 5 [Link] (https://github.com/dlab-berkeley/DIGHUM101-2020/tree/master/Notebooks/Week5).
Dave, Pranay. "PCA vs TSNE - ElClásico." Medium, Towards Data Science, 30 May 2020,
[Link](towardsdatascience.com/pca-vs-tsne-el-cl%C3%Alsico-9948181a5f87.)
Tran,Khuyen. "How to Solve Analogies with Word2Vec." Medium, Towards Data Science, 29 Mar. 2020,
[Link](towardsdatascience.com/how-to-solve-analogies-with-word2vec-6ebaf2354009)
Kulshrestha, Ria. "NLP 101: Word2Vec - Skip-Gram and CBOW." Medium, Towards Data Science, 19 June
2020,[Link](towardsdatascience.com/nlp-101-word2vec-skip-gram-and-cbow-93512ee24314).
Amipara, Kevin. "Better Visualization of Pie Charts byMatPlotLib." Medium, Medium, 20 Nov. 2019,
[Link](medium.com/@kvnamipara/a-better-visualisation-of-pie-charts-by-matplotlib-935b7667d77f).
Li,Zhi. "A Beginner's Guide to Word Embedding withGensimWord2Vec Model." Medium, Towards Data Science, 1 June
2019,[Link](towardsdatascience.com/a-beginners-guide-to-word-embedding-with-gensim-word2vec-model-5970fa56cc92).
Scott, William. "TF-IDF for Document Ranking from Scratch in Python on Real World Dataset." Medium, Towards Data Science, 21 May 2019,
[Link](towardsdatascience.com/tf-idf-for-document-ranking-from-scratch-in-python-on-real-world-dataset-796d339a4089).
Brems, Matt. "A One-Stop Shop for Principal Component Analysis." Medium, Towards Data Science, 10 June 2019,
[Link](towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c).
jeffd23. "Visualizing Word Vectors with t-SNE." Kaggle, Kaggle, 18 Mar. 2017, [Link](www.kaggle.com/jeffd23/visualizing-word-vectors-with-t-sne).
alvasalvas, et al. "Ensure theGensimGenerate the Same Word2Vec Model for Different Runs on the Same Data." Stack Overflow, 1 Aug.
1965,[Link](stackoverflow.com/questions/34831551/ensure-the-gensim-generate-the-same-word2vec-model-for-different-runs-on-the-sam).
Gaudard, Olivier. "#11 GroupedBarplot." The Python Graph Gallery, 7 Sept. 2017, [Link](python-graph-gallery.com/11-grouped-barplot/).
RaulGonzalesRaulGonzales et al. "CantPlot Seaborn Graphs Side by Side." Stack Overflow, 1 May 1969,
[Link](stackoverflow.com/questions/58624647/cant-plot-seaborn-graphs-side-by-side).
DreamsDreamsand NielsJoaquinNielsJoaquin 1. "Visualiseword2vec Generated fromGensim." Stack Overflow, 1 Dec. 1966,
[Link](stackoverflow.com/questions/43776572/visualise-word2vec-generated-from-gensim).
oceandyeoceandyeandCathyQianCathyQian. "How Do I Save Word Cloud as .Pngin Python?" Stack Overflow, 1 Apr. 1968,
[Link](stackoverflow.com/questions/52464932/how-do-i-save-word-cloud-as-png-in-python).
RobertFrankeRobertFranke et al. "Save a Subplot in Matplotlib." Stack Overflow, 1 June 1960,
[Link](stackoverflow.com/questions/43257733/save-a-subplot-in-matplotlib).
Zagorax, et al. "What Is the Operation behind the Word Analogy in Word2vec?" Stack Overflow, 1 Apr. 1968,
[Link](stackoverflow.com/questions/52364632/what-is-the-operation-behind-the-word-analogy-in-word2vec).
alvasalvas"Interpreting Negative Word2Vec Similarity fromGensim." Stack Overflow, 1 Sept. 1966,
[Link](stackoverflow.com/questions/42381902/interpreting-negative-word2vec-similarity-from-gensim).
ShaktiShakti"How to RunTsneon word2vec Created fromGensim?" Stack Overflow, 1 June
1966,[Link](stackoverflow.com/questions/40581010/how-to-run-tsne-on-word2vec-created-from-gensim).