Society of Women in Engineering (SWE) Natural Language Analysis of Outreach Data

Agnes Priyadharshini Antony Joseph
Department of Computer Science
Texas Tech University
Lubbock
agnes-priyadharshini.antony-joseph@ttu.edu

Abstract— **The main objective of this paper is to analyze the good practices in events conducted by the Society of Women in Engineering (SWE). In order to make the events more enjoyable and interactive, it is not only necessary to study the list of activities performed at each event, but also study which activity had a positive impact on the participants. In other words, activities conducted in every SWE event brought them closer to the field of Engineering. In addition, it is important to analyze the participant's views on SWE events. These events and views are the two main factors that contribute towards the success of an event. In the analysis, TextBlob, Term Frequency Inverse Document Frequency (TF-IDF), SpaCy, Jaccard Similarity, Latent Dirichlet Allocation (LDA) and the Multinomial Naïve Bayes Classifier are used. TextBlob is used to perform sentiment analysis. It is observed that 75% of the comments are positive. Further, TF-IDF and SpaCy are implemented to extract strips of information about the activities performed in each event. For the sake of extracting unique activities, Jaccard Similarity is executed. Also, Latent Dirichlet Allocation (LDA) is performed to extract the topics discussed by the participants. In the interest of exploring the suggestions given by the participants, Multinomial Naïve Bayes Classifier is used.**

Keywords— SWE, Engineering, TextBlob, TF-IDF, SpaCy, LDA, Jaccard Similarity, Multinomial Naïve Bayes Classifier

I. Introduction

The Society of Women in Engineering is a national non – profit organization. The main goal of SWE is to help women in their professional journey as engineers. SWE accomplishes its goal by supporting women financially right from their college days by awarding various scholarships. It motivates women belonging to different levels by conducting numerous outreach events throughout the year thus providing them an opportunity for leadership, mentoring and volunteering.

By the end of each event, suggestions are collected from the participants by the SWE Outreach Committee using the Outreach Metric Tool (OMT) so as to track the characteristics of the event. Normally the comments received can be classified into three types based on the sentiment namely positive, negative and neutral. Neutral comments are instructions more likely to be followed by the participants during the event; for example, please read the pamphlets for more details. Positive comments describe how the participants benefited from the event. Negative comments focus on the difficulties faced by the participants and the kind of activities that the participants disliked (or found boring) during the event. Analyzing such comments helps us draw insights which in turn contribute towards the success of future events.

II. Literature Review

According to [1], narrow range of people choose STEM careers. The reason why less number of people choose STEM careers is because of the lack of awareness or education about STEM. In order to educate the students about STEM, various outreach events are conducted. The main focus of the outreach events is to educate students who are in their transition points, for example, high school to college. In short, STEM career awareness is given to pre college students. The main aim of the outreach events is to promote the awareness of engineering profession, provide academic enrichment, have mentors trained in the field of engineering who will act as role models for the students to pursue a career in STEM and to be supported by the educational system of the student participants. By [1], educating students have increased the percentage of students who chose science, technology, engineering and mathematics as their career fields. Also, educating undergraduates belonging to a different field will help them opt for a dual degree engineering program. Further, help them pursue advanced degrees and research careers in the field of science and engineering.

[2] researches about the difficulties faced by female students while getting exposed to STEM. The research is done based on three main categories. In childhood and adolescence, girls do not have a personal goal and have little knowledge

about STEM. In emerging adulthood, due to out numbered male peers, female students do not get to know about any female role models in the field of STEM. Hence, girls leave STEM field prematurely. In early to mid adulthood, gender bias in hiring and promotion and biased judging of work undermine the retention of women in STEM. Further, [2] has also concluded that evidence based programs like K 12 outreach events, after school activities, exhibitions and mentoring groups can help girls know more about engineering.

As mentioned in [3], Washington DC based commission of Professionals in Science and Technology and Engineering have found that lowest percentage of women are found in STEM among all professions. The main reason for the lowest percentage of women is that over three quarters of the women do not know what engineering is and what engineers do. In short, engineering accomplishments are invisible. In addition, [3] introduces Jill' theory which states that companies have stopped touting their achievements and that technology has gone underground. Though engineering lays a major role in our everyday routine, the practitioners have lost their visibility and allure. The solution to this problem, as described in [3], is to make engineering visible by conducting many outreach events.

In [4], the authors talk about the gender disparities in the field of engineering. In order to reach young women and engage them in engineering, outreach events are conducted. In particular, Females Excelling More in Math Engineering and Science (FEMMES) is a free event organized to attract 4th - 6th grades towards the field of engineering by conducting many hand on activities. Furthermore, for the success of FEMMES events in future, feedback is collected from the participants before and after the event. After a thorough analysis on the comments, the results clearly indicate that students enjoyed activities related to the field of Engineering than Maths and Science.

In order to increase the interest of pre university students towards the field of Science and Engineering, the authors in [5] state that IEEE has conducted a special session that aims to discuss the impacts of the programs conducted over the past 10 years. As the result, they have discovered many problems and provided out of the box solutions to all the problems.

Due to the increasing demand for STEM professionals, a STEM based outreach program has been established to inspire young minds. In [6] these STEM programs have been analyzed in such a way that the STEM outreach events will have a long term impact on the participants. The main goal of this paper is to identify the different age ranges and the type of activities that inspire or have a long term effect by age. The paper clearly identifies that from preschool till 4 years of age, STEM and engineering concepts can be introduced to the kids, From 5 - 6 years, the students can start learning about STEM and engineering and from 7 years - 12 years the students can get involved in lots of hands on activities. This age range identification will help us decide on the type of activities to decide based on the age range of the participants.

III. Dataset Description

The dataset consists of details about each event conducted by SWE from 2014 – 2018. The dataset contains the following information,

Feature	**Description**
City/ Town	City/ Town where the event was conducted
State	State in which the event was conducted
Country	Country in which the event was conducted
Duration of the Event	Number of hours the event lasted
Type of event	Description is given based on the activities performed in the event, eg: discussion, outreach event, hands on event, etc…
Number of girls under 18 years	Number of female participants who were under 18 years of age
Number of boys under 18 years	Number of male participants who were under 18 years of age
5 – 11 years old	Number of participants who belonged to Elementary School
11 – 14 years old	Number of participants who belonged to Middle School
14 – 18 years old	Number of participants who belonged to High School
Other	Number of participants who were 18+
Number of Adult Guests	Adults could be parents, Educators or any non – volunteers who were a part of the event
Number of SWE volunteers	Includes SWE members who were volunteers at the event and who helped in planning the event. Even

	though a few of them did not participate in the event, they are considered to be a SWE volunteer
Number of other volunteers	Includes non - SWE members who were volunteers at the event and who helped in planning the event. Even though a few of them did not participate in the event, they are considered to be a non - SWE volunteer
Did a Partner Organization Participate?	Indicates if the event was conducted by SWE or if the event was conducted by SWE and any partner organization, example, Girls Scouts
We received a SWE Program Development Grant (PDG) for this event	Denotes if the event has received a PDG grant
Please share any best practices, lessons learned, additional description, or comments regarding your Event or Activity.	Contains all the comments given by the participants about the events

IV. Research Question

Analyzing what participants feel about the events conducted by SWE will help us yield insights like the activities enjoyed by the participants, top universities/schools that get involved in SWE events, which particular event helped students to pursue a career in engineering, and to figure out the suggestions given by the participants. Our major research focuses on performing sentiment analysis, topic modelling and summarization on the comments provided by the participants through the OMT.

V. Methods

A. Text Blob

Text Blob is a Python library for processing textual data. It performs natural language tasks, such as identifying the parts of speech tagging, noun phrase extraction, sentiment analysis, classification, translation and many more. In this paper, Text Blob is used to perform sentiment analysis on the comments given by the participants about SWE events. Text Blob identifies, quantifies, extracts and studies affective states and subjective information from the comments and uses this information to identify the comment as positive, negative or neutral.

The sentiment of a comment is identified by computing it' s polarity [7]. Usually, the value of sentiment polarity varies between -1.0 to 1.0. The comments with a positive polarity are considered to be positive comments, comments with a negative polarity are classified as negative comments and the comments with a polarity value of 0.0 are categorized as neutral comments. In order to extract good practices in SWE events and explore the areas to be improved positive and negative comments are used in further steps.

B. Term Frequency – Inverse Document Frequency (TF–IDF)

TF-IDF is a numerical statistic that determines how important a word is to a document. The number of times a word appears in a document is known as term frequency. Term frequency tf(t,d) is defined as the raw count of the term t in the document d [8][9]. The formula to calculate term frequency is as follows,

$$\mathrm{tf}(t,d) = 0.5 + 0.5 \cdot \frac{f_{t,d}}{\max\{f_{t',d} : t' \in d\}}$$

where,

t denotes the term

d denotes the document

$f_{t,d}$ denotes the frequency of term t in the document d

Normally, stop words have a higher term frequency. The words that are most commonly used in a language are termed as stop words. Hence stop words are removed using NLTK.

With term frequency all the terms are considered to be equally important. Some words except the stop words may occur commonly; for example, 'that', this word is not an important term in the document. In order to minimize the weight of common terms inverse document frequency is computed.

Inverse document frequency is defined as logarithmically scaled inverse fraction of documents that contains the term divided by the total number of documents. The formula to calculate inverse term frequency is as follows [10][11],

$$IDF(w_i) = \log\left(\frac{|D|}{DF(w_i)}\right)$$

$w_i = i^{th}$ term

d = document, d (i) = TFIDF of term w_i in document d

$TF(w_i, d)$ = Term Frequency of term wi in document d

$IDF(w_i)$ = Inverse Document Frequency

In furtherance of figuring out how important a term is to a document, TF-IDF is computed. TF–IDF is determined using the formula below[10][11],

$$d^{(i)} = TF(w_i, d).IDF(w_i)$$

In short, the TF–IDF score closer to 1 denotes that the term rarely appears in the document, whereas, if the TF–IDF score is close to 0, it denotes that the term appears quite often in the document.

In this paper, the document is split into chunks (ngrams) and TF-IDF score is calculated for each chunk so as to extract important information.

C. SpaCy

SpaCy is an open source library for advanced natural language processing. It is written with two programming languages namely Python and Cython. A few of the main features of SpaCy are non - destructive tokenization, named entity recognition, parts of speech tagging, labelled dependency tagging and text classification [18].

In this paper, SpaCy is used to recognize the major universities and schools that participate in SWE events using named entity recognition. In addition, parts of speech tagging is implemented on the chunks of sentences with a higher TF-IDF score and these sentences are matched for patterns that identify the sentences that enclose information related to the activities performed in the SWE events.

D. Jaccard Similarity

The sentences that encloses information about SWE activities may have similar context. In order to extract unique list of activities performed in SWE events, jaccard similarity is used. Jaccard similarity is defined as the size of intersection between the two sets divided by the union of two sets [12]. The formula to calculate Jaccard similarity is defined below,

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

where,

A and B are two sets

As defined above, A is the set that contains words from the first comment and B is the set that contains words from the second comment. This paper ignores all the activities that are more than 30% similar.

E. Latent Dirichlet Allocation (LDA)

Each document is composed of several topics. To identify what the document talks about and all the important topics that the document covers, it is necessary to read the entire document. Manual identification of distributed topics is possible if the size of the document is small but as the size of the document increases, the probability of reading the entire document decreases eventually. In such a scenario, LDA can be applied to extract the mixture of topics in the form of words with their respective probabilities [13][14].

This paper uses LDA to analyze the different areas on which participants comment. Identifying different topics in turn helps us identify the area that we have to focus on for the success of future SWE events.

F. Summarizer

Gensim is a Python library modelled to handle larger text collections using data streaming and efficient incremental algorithms. When given a huge document, it is always preferred to read the document's summary than the entire document. In Gensim, an inbuilt summarizer summarizes the entire document based on the Text Rank Algorithm.

Text Rank Algorithm is similar to the Page Rank Algorithm. Page Rank Algorithm works by counting the number of quality links to a page to determine an estimate of how important the website of the page is [15]. It is built on the assumption that more important websites are likely to receive more links from other websites. Text Rank Algorithm works with the same principle but web pages are replaced by sentences in the document [16]. In case of huge documents, a number between 0 and 1 can be specified so as to determine the percentage to which the entire document has to be summarized. By default, Gensim's summarizer briefs the entire document to 20%. Likewise, a keyword summary can be constructed using keywords from Gensim. In this case, the sentences are replaced by words. Time taken to build a document summary grows as the size of the document increases. This paper uses the Gensim summarizer to build a summary on feedback provided by the participants. Additionally, the Gensim keyword summarizer is used to build a keyword summary so as to identify the important terms used.

G. Multinomial Naive Bayes Classifier

When important sentences from a document are extracted using TF-IDF as mentioned above, to separate the suggestions from good practices Multinomial Naive Bayes Classifier is used [17].

A training set is built using the Yelp reviews dataset. Based on the rating given by each user to a restaurant, the review posted by the user is collected and classified into suggestions (if the rating is below 3) and good practices (if the rating is above 4). Using this training set and Multinomial Naive Bayes Classifier, a model that identifies suggestions and good practices is built.

VI. Results

A. State based analysis

The graph below represents the number of positive, negative and neutral comments provided by the participants

from top 8 states that belong to the United States. In Figure 1, state based sentiment analysis, depicts that all the top eight states of United States have good amount of positive review but California (CA), Illinois (IL) and New York (NY) have equal amounts of negative reviews.

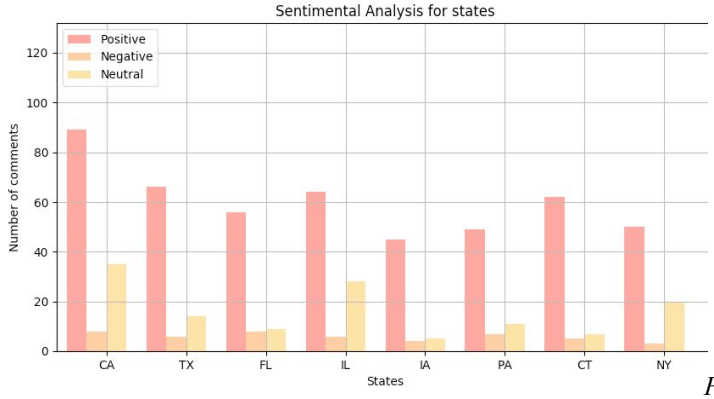

Figure 1. State based sentiment analysis

Top universities and schools that participate in SWE events are identified from the comments using name entity recognition from SpaCy. Figure 2.1 and 2.2 represent the universities and schools that participated in SWE events by each state and clearly portray that California, Texas, Pennsylvania and Connecticut have a good amount of universities and schools that participate in SWE events whereas Florida and Iowa have only one frequently participating organization.

California	**Texas**
Bourns College of Engineering Santa Clara University Atascadero Middle School Templeton High School HS & College College of ECC Santa Maria Middle Middle school San Luis Obispo High School San Jose State University College of Engineering	The Ann Richards School Coronado High School Clint Small Middle School Denton High School the University of Texas the Cockrell School Bryan Adams High School University of Houston
Florida	**Illinois**
Destin Middle School	University of IL Extension Bradley University the University of Illinois at Chicago

Figure 2.1. Universities/ Schools that participate in SWE events

Iowa	**Pennsylvania**
The University of Northern Iowa	Lehigh University SWE Widener University Ohio State University Independence Charter School the College of Engineering College of Engineering
Connecticut	**New York**
John F. Kennedy School Norfeldt School Betances STEM School the Central Connecticut State University Wesleyan University Public Schools Collaborative Plainville High School	Acadia Middle School Students Stony Brook University the Robert C. Parker school

Figure 2.2. Universities/ Schools that participate in SWE events

Keywords used in each state are represented in the form of a word cloud. Keyword summarization is done using Gensim. Figure 3 represents keyword summary for California,

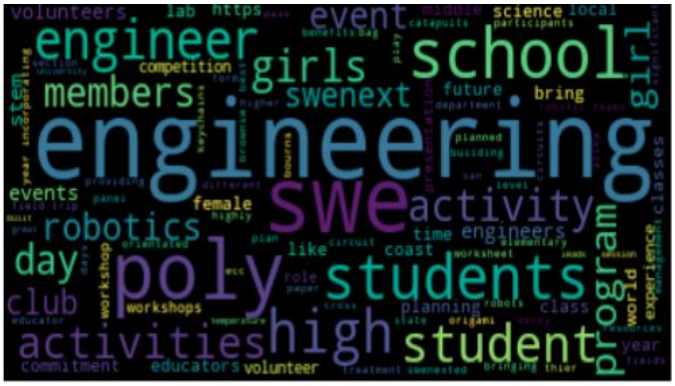

Figure 3. Keyword summary of comments given by participants about the events conducted in California

From Figure 3, it is identified that activities such as building structures, discussing about SWENext and presentations on which career to choose in the future are the major activities held in California.

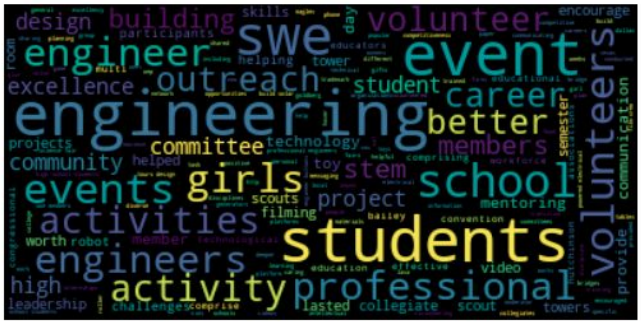

Figure 4. Keyword summary of comments given by participants about the events conducted in Texas

Figure 4 represents that many outreach events were conducted and technology related activities were performed by the students in the state of Texas. Also, events in Texas serves as a place to grow the quality of leadership. In case of outreach events, the enrollment has been improved.

Figure 5 exemplifies the different types of activities performed in Florida such as organizing workshops, building robots, making brownies, water related activities, activities that includes magnets, reading books and learning the idea behind floating.

Figure 6 exhibits the type of activities performed in Illinois which includes chemical reactions, conducting interviews, recreation activities, discussion about the newsletter and special Halloween activities. In addition, interaction events are held between the mentors and the students during the lunch period.

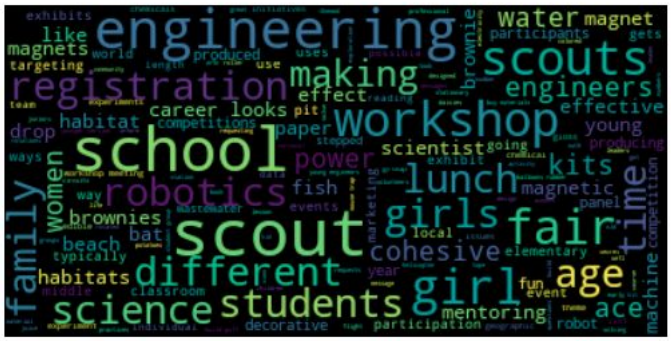

Figure 5. Keyword summary of comments given by participants about the events conducted in Florida

Figure 6. Keyword summary of comments given by participants about the events conducted in Illinois

Figure 7. Keyword summary of comments given by participants about the events conducted in Iowa

In Figure 7, activities like building cars, parachutes, building simulators and chemical reactions were organized. It is evident that many kids were involved in these SWE events held at Iowa.

Figure 8 denotes that the SWE events were heavily publicized and that parents were also involved in all the activities in the state of Pennsylvania.

Figure 9 represents that many activities that involved paper were organized and professional events were held at the state of Connecticut.

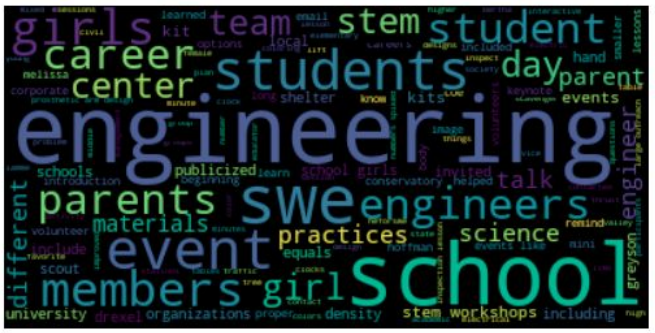

Figure 8. Keyword summary of comments given by participants about the events conducted in Pennsylvania

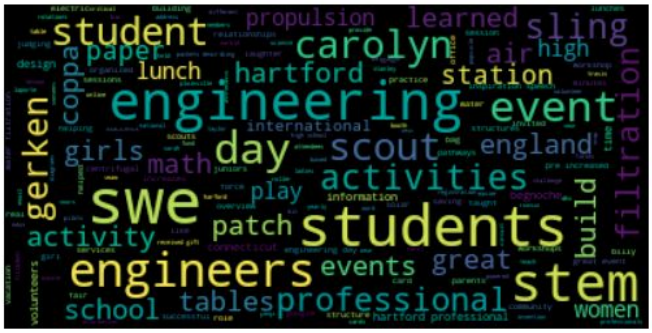

Figure 9. Keyword summary of comments given by participants about the events conducted in Connecticut

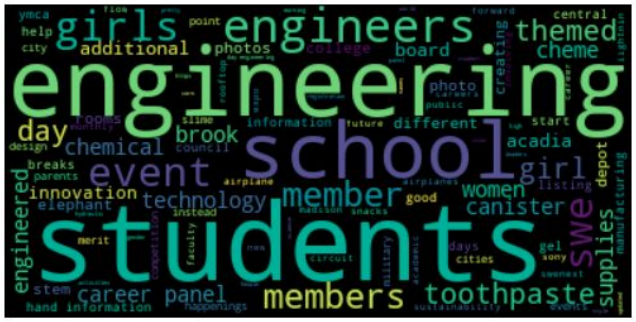

Figure 10. Keyword summary of comments given by participants about the events conducted in Connecticut

Figure 10 depicts that the list of events held in Connecticut includes building structures, building slingshots, paper related activities including art and craft and learning the ideas behind propulsion. In addition, Girls Scouts were the major organization that partnered with SWE in Connecticut.

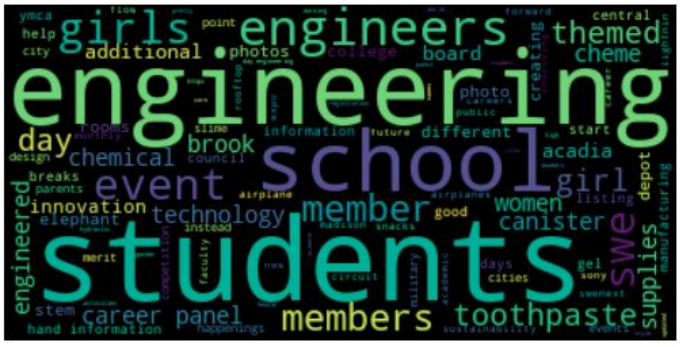

Figure 11. Keyword summary of comments given by participants about the events conducted in New York

Figure 11 depicts that the list of activities performed in New York includes technology related activities, making toothpaste, presentations on which career path to choose and experimenting with chemicals. Further, parents were involved in these events.

Using LDA, the top 5 topics discussed by the participants in their comments are extracted. Figure 12, 13, 14, 15, 16, 17, 18 and 19 represents the topics discussed in the comments given by the participants about SWE events in California, Texas, Florida, Illinois, Pennsylvania, Connecticut, New York and Iowa respectively. These topics talk about workshops, presentations, lectures and activities performed during SWE events. Further, discussions have been performed on PDG report.

Topic 1	Topic 2	Topic 3
Event	Present	FIRST
Differ	Engineering	Get
Science	Work	Organization
Carnival	Gave	Engaging
Halloweentheme	San	Learn
Stick	Swenext	Year
Craft	Keep	Delegation
Good	Kid	Done
Want	Fun	Load
STEM	various	play

Figure 12. Topic wise distribution of words in California.

Topic 1	Topic 2	Topic 3
Cream	Competition	Learn
Ice	Catapult	Chemistry
Number	Portion	Student
Denton	Build	Involve
High	Inc	Get
Estimate	Love	Help
Girlstart	Event	General
Girls	Girl	Concept
School	Include	Bag
bag	robot	construct

Figure 13. Topic wise distribution of words in Texas

Topic 1	Topic 2	Topic 3
Great	Success	Grader
Checklist	Key	Older
Family	Science	5th
Engineering	Checklist	answer
Event	Judge	Girl
Organization	Fair	Question
Fun	Great	discuss
Control	Get	book
Software	School	Read
day	smaller	get

Figure 14. Topic wise distribution of words in Florida

Topic 1	Topic 2	Topic 3
Event	Girl	Ahead
Workshop	Survey	Time
Like	Science	Demonstration
Post	Preworkshop	Logistic
Response	Train	Handson
Member	High	Transport
Easy	Workshop	Video
Survey	School	Booth
Kid	Fair	Schedule
lecture	experience	plan

Figure 15. Topic wise distribution of words in Illinois

Topic 1	Topic 2	Topic 3
Two	Organization	Girl
Plus	Student	Kid
Kid	Hands on	angelika
Winner	Engineering	Material
Interest	Differ	Behaviour
Severs	Speaker	Need
College	School	Physical
Open	Women	Industry
Teacher	Keynote	Pipe
patent	high	foundation

Figure 16. Topic wise distribution of words in Pennsylvania

Topic 1	Topic 2	Topic 3
Coaster	Puff	System
Roller	Build	Filtration
Water	Time	3rd
Boat	Mobile	Improve
Mobile	Wonder	Water
Penny	List	Activity
Filtration	Thing	Through
Shot	Penny	Well
Air	Make	Great
hoopster	scout	5th

Figure 17. Topic wise distribution of words in Connecticut

Topic 1	Topic 2	Topic 3
School	Engineering	Member
Event	Swe	Made
Booth	Explain	Help
High	Question	Interest
Parent	Experience	Career
Demonstration	Rpi	Participate
Year	Host	Present
Dinner	Answer	Start
Would	Use	Inform
girl	college	swe

Figure 18. Topic wise distribution of words in New York

Topic 1	Topic 2	Topic 3
Student	Larger	Make
Always	Group	Bottle
Help	Worst	Cone
Love	Prepare	paper
Science	Expectation	Launch
Great	Back	Test
Experiment	Well	Enjoy
Fun	Work	Earlier
well	Sure	Multiple
done	activity	prepare

Figure 19. Topic wise distribution of words in Iowa

B. Analysis based on Program Development Grant (PDG)

ExxonMobil Foundation supports SWE activities by aiding them financially. This financial micro-grant is termed as the PDG [25]. Figure 20 represents the top events that receive PDG out of which hands on activities and mentoring groups have received most of the grants,

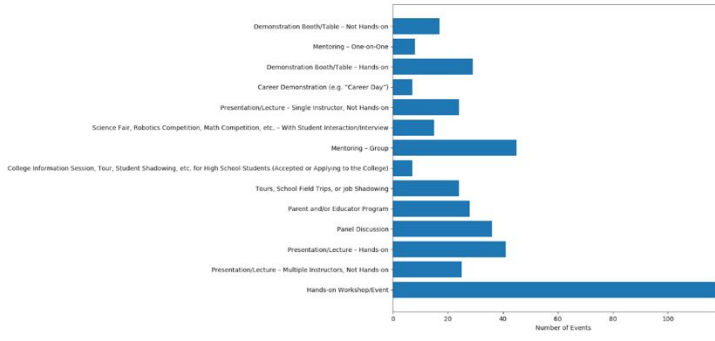

Figure 20. Top events that received PDG

Not all the events receive a PDG. Figure 21 represents the top events organized by SWE that do not receive PDG which represents that many hands on activities, presentation lectures and demonstration booth tables have not received the grant,

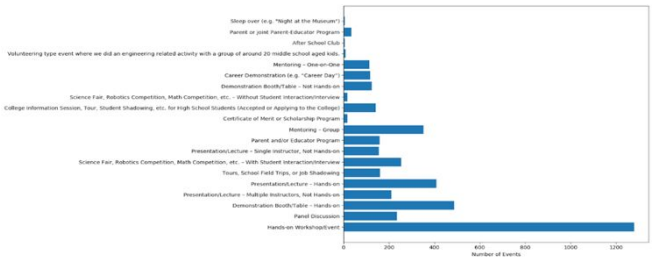

Figure 21. Top events that do not receive PDG

Mostly when a PDG is sanctioned for an event, it increases the quality of the event. For example, the location where the event is organized will be able to support a big crowd or buses could be provided to reach the destination, etc., thereby reducing the possibility of getting a negative comment from the participant.

On the contrary, when PDG is not received, this reason could serve as a possibility for the participants to give a negative feedback about the event. Since it is highly unlikely to receive a PDG grant on every event, top activities enjoyed by the participants when a PDG is not received are studied using pattern matching. Figure 22 represents the activities enjoyed by participants,

```
built balloon towers so that
built catapults using limited suppli
run tables to explain the
helped run tables to explain
run tables to explain the
Using Snap Circuits kits only
were hands on and for
made ice cream in a
planning something the same day
represent Mechanical Engineering at
attend college  Several other
had patents pending on their
Gave students edible citric acid
baking soda and jello powder
was POOR but the competition
created paper airplanes out of
built paper towers The group
```

Figure 22. Activities enjoyed by the participants – SWE events without PDG

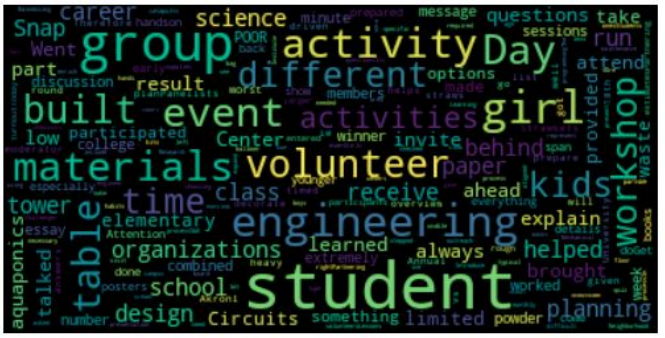

Figure 23. Keyword summary of activities enjoyed by the participants – SWE events without PDG

Figure 23 represents the list of activities enjoyed by the participants when the PDG is not received. Major activities include building tower, experimenting with snap circuits, workshops, learning the ideas behind aquaponics, competitions and presentations.

C. Analysis based on Partnership with Other Organizations

SWE organizes many events to bring people closer to engineering. Figure 24 represents the events organized by SWE without any partners, such as many hands on activities, demonstration booth tables and mentoring groups have been organized,.

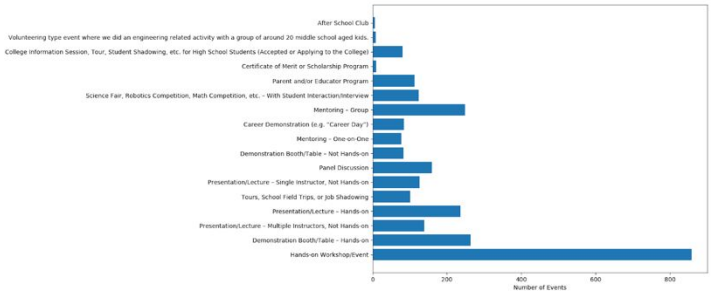

Figure 24. Events organized without partners

Generally, when any event is organized, the number of volunteers available to support the event plays an important role to make the event a great success. Figure 25 represents the list of number of events organized by SWE with partners, many hands on activities, demonstration booth tables and mentoring groups are organized.

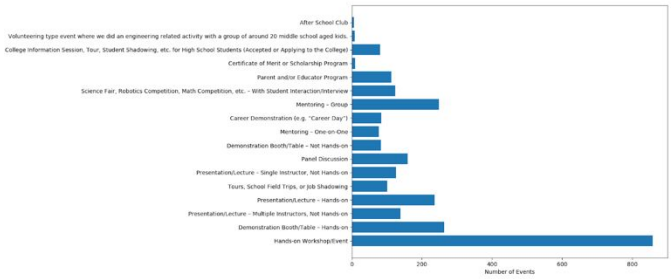

Figure 25. Events organized with partners

Figure 26 represents the top organizations that collaborate with SWE out of which Girls Scouts and FIRST partner with SWE,

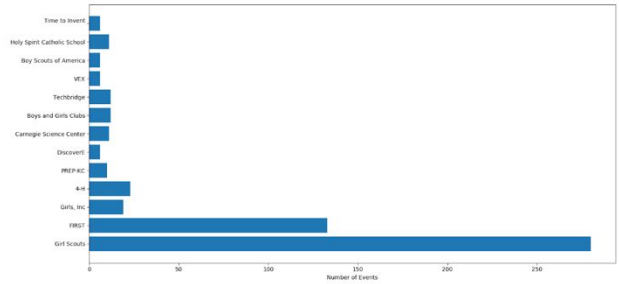

Figure 26. Top organizations that partner with SWE

A huge difference when you team up with another organization is that the number of resources allocated for the event expands. To assist the participants throughout the event, both SWE and non-SWE volunteers are available. From Figure 26, it is identified that SWE partners with Girls Scouts and FIRST organizations quite often.

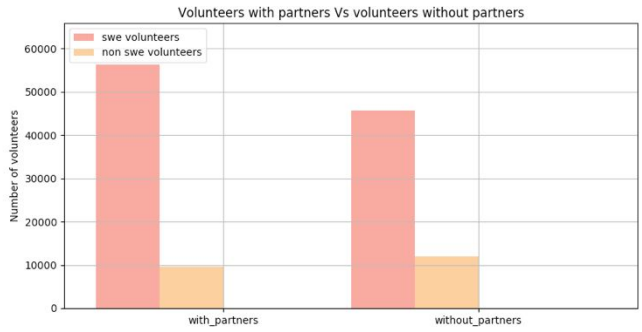

Figure 27. Difference in number of volunteers when an organization is partnered

Figure 27 represents the number of SWE and non SWE volunteers when SWE teams up with an organization versus SWE and non SWE volunteers when there are no partners.

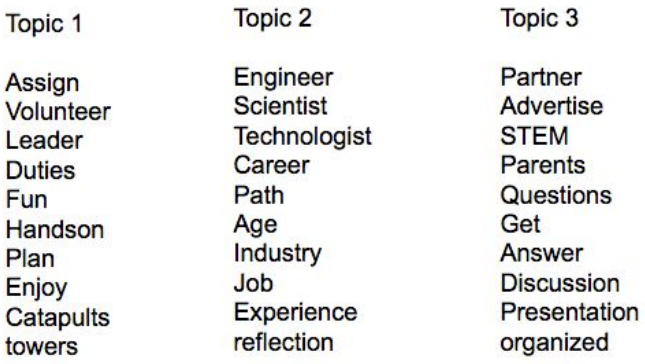

Figure 28. Topic analysis on comments given by the participants on the events with partner organizations

Figure 28 represents the top three topics discussed by the participants on the events with partner organizations. The topics include discussions about STEM , handson activities and professional experiences as engineers.

Figure 29 depicts the type of events organized when SWE collaborates with Girls Scouts out of which lot of events are workshops, lectures, educator programs and mentoring groups,

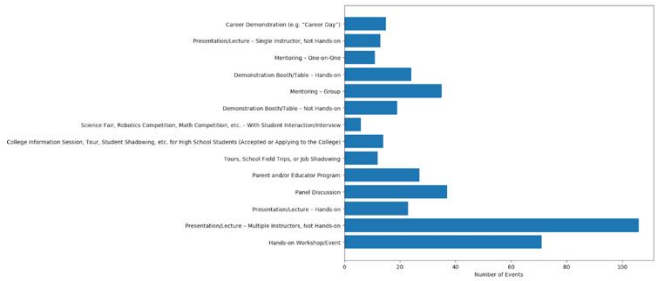

Figure 29. Events organized with Girls Scouts

Fig 30 represents the top activities enjoyed by the participants when Girls Scouts are involved. Major activities include experimenting with soldering kits, building roller coasters with paper, filtration systems, flashlights, slingshots, bridges with popsicles, experimenting with galaxy slime and learning the ideas behind floating.

Figure 30. Activities enjoyed with Girls Scouts

Figure 31 represents the top events organized by SWE when they acquire assistance from the FIRST organization. The events include presentations, lectures, hands on and not hands on workshops and demonstration booth tables.

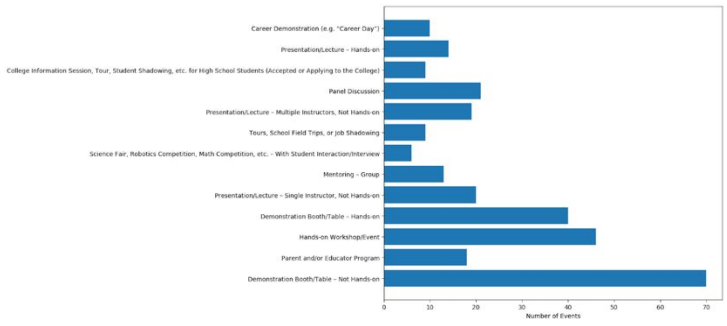

Figure 31. Events organized with FIRST

Analyzing the comments given by the participants when SWE partners with FIRST reveals that mentoring, building robots and lectures on STEM, are liked by the participants. In addition, the practice wear provided by FIRST is also liked by the participants.

D. Analysis based on Outreach Events

SWE members are determined to inspire women to become engineers [25]. SWE keeps working towards its goal by conducting numerous outreach events. Since spreading the goodness of engineering and bringing women closer to engineering is SWE's main goal, they conduct numerous outreach events to reach the students. Figure 32 represents the number of outreach events conducted every year,

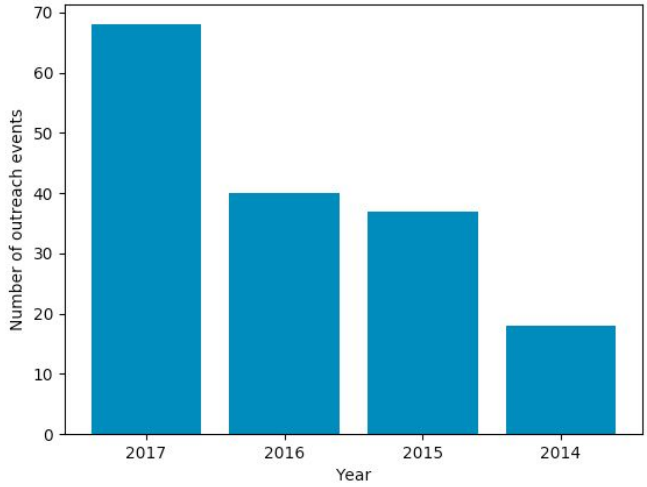

Figure 32. Number of outreach events conducted every year from 2014 -2017

From Figure 32, it is clearly identified that SWE has a quick progress in working towards its goal. To make the outcome of these events efficacious, comments are collected from the participants after each event. Figure 33 represents sentiment analysis on these comments.

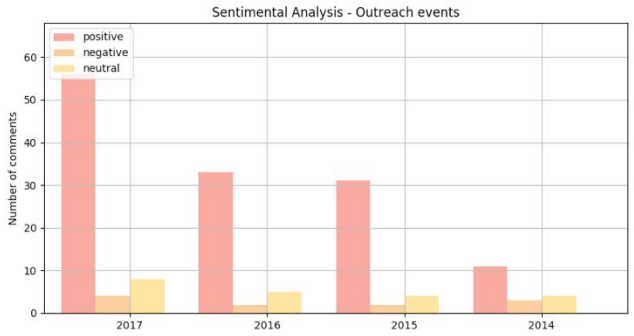

Figure 33. Sentimental analysis on the comments received during outreach events

Figure 33 reveals that in addition to the increase in number of outreach events conducted every year, the positive reviews from the participants have also increased.

To receive a productive output, suggestions are extracted from the comments given by the participants. The top suggestions are a request to increase the time of the lectures and activities. Also, more focus is required in the areas of hands on activities.

VII. Future work

Analyzing the best practices and suggestions can be done for every event so as to identify the activities enjoyed by the participants under both hands on and not hands on, to identify lectures given in which particular events have motivated students to choose engineering as their careers, and to identify the person whose lectures or presentations are motivating. In addition, replace Jaccard similarity with Minhash so as to improve the time complexity when the data grows.

VIII. Conclusion

The events organized by SWE to circulate the notion of engineering among women is analyzed in this paper. As the result of state based evaluation, it is detected that SWE events are concentrated in the states of California, Texas, Florida, Illinois, Iowa, Pennsylvania, Connecticut, and New York. The top universities and schools that participate in SWE events are less in comparison with the list of universities and schools available in each state. Further, SWE events have a huge demand.

Analysis based on PDG serves as a proof that many events do not receive a PDG due to limited PDG funding.

Analysis based on partnership with other organizations proves to have a large number of volunteers when SWE teams

up with another organization. Further analyzing the comments that involve Girls Scouts proves that Girls Scouts team up with SWE for many hands on events and a large number of participants enjoy the activities with Girls Scout. In addition, inspecting the comments that involve FIRST depicts that FIRST is involved in a variety of events namely mentoring, discussions, hands on, and not hands on activities. Also, it is evident that the participants enjoy FIRST Mentoring activities.

While examining the outreach events conducted by SWE, it is proved that the number of outreach events has progressed since 2014. The achievement is that the number of positive results have also increased every year along with the number of outreach events.

To sum up, as the number of SWE events increase every year, the percentage of positive comments tend to increase and the percentage of negative comments are less than 15 %. It is evident from the results that PDG does not play a major role in deciding the success of a SWE event. Furthermore, top activities enjoyed by the participants include building structures and performing science related experiments. Also, presentations and lectures on STEM and engineering seems to have created a huge impact on all the participants.

References

[1] Chubin, Daryl E. , May, Gary S. , Babco, Eleanor L. ., Diversifying the Workforce , Journal of Engineering Education, January 2005, Vol. 94(1), pp. 73-86.

[2] Nilanjana Dasgupta, Jane G. Stout ., Girls and Women in Science, Technology, Engineering, and Mathematics: STEMing the Tide and Broadening Participation in STEM Careers, Policy Insights from the Behavioral and Brain Sciences 2014, Vol. 1(1) , pp. 21–29.

[3] Jill S. ., Why so few Women Still? IEEE Spectrum, 2004, Vol. 41, Issue. 10.

[4] Victoria W.,Aubrey B.Allison E.,Heileen H.,Gary Y.,Working in Progress:A STEM educational Outreach Day for young Females. Frontiers in Education Conference, 2008.

[5] Sohaib Qamar S.,Efthymia A., STEM education outreach through IEEE's pre-university programs — Engaging volunteers to benefit K-12 education and local communities, 2015.

[6] Mr. Ralph T., Dr. Mo M.,Influencing Factors in Identifying the Optimal Age for Conducting STEM Outreach.Integrated STEM Education Conference (ISEC), 2016.

[7] Akhilesh Kumar Singh, Deepak Kumar Gupta, Raj Mohan Singh ., Sentiment Analysis of Twitter User Data on Punjab Legislative Assembly Election, 2017, International Journal of Modern Education and Computer Science, Vol. 9, pp. 60-68.

[8] Kranti Ghag, Ketan Shah ., SentiTFIDF – Sentiment Classification using Relative Term Frequency Inverse Document Frequency, International Journal of Advanced Computer Science and Applications, 2014, Vol. 5, No. 2.

[9] Tao Peng, Lu Liu, Wanli Zuo ., PU text classification enhanced by term frequency–inverse document frequency-improved weighting, Concurrency and computation: Practice and Experience, 2014, Vol 26, Issue 3.

[10] G. Salton and M. McGill, "Introduction to modern Information Retrieval," McGraw-Hill, 1983, pp. 105-107 & 205,.

[11] J. Han and M. Kamber, "Data Mining Concepts and Techniques," Morgan Kaufmann Publishers, 2006, 02nd edition, pp. 364-365.

[12] Kato, T., Shimizu, I. & Pajdla, T. ., Selecting image pairs for SfM by introducing Jaccard Similarity, 2017, IPSJ T Comput Vis Appl Vol. 9 Issue. 12.

[13] YueGuo, Stuart J.Barnes, Qiong Jiaa ., Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation, 2017, Tourism Management, Vol. 59, pp. 467-483.

[14] Brian P. Eddy, Nicholas A. Kraft, Jeff Gray, Impact of structural weighting on a latent Dirichlet allocation–based feature location technique, 2018, Journal of Software Evolution and Process, Vol. 30, Issue. 1.

[15] Maria Pershina, Yifan He, Ralph Grishman ., Personalized Page Rank for Named Entity Disambiguation, 2015, Human Language Technologies: The Annual Conference of the North American Chapter of the ACL, pp. 238–243.

[16] Chuanqi Lan, Hongzhi Yu, Tao Xu, Peixin Liu, Jiuyi Li ., Research on keyword extraction of Tibetan web news based on improved TEXT-RANK algorithm, 2017, Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE 2nd Information.

[17] Pintu Lohar, Koel Dutta Chowdhury, Haithem Afli, Mohammed Hasanuzzaman, Andy Way ., ADAPT at IJCNLP-2017 Task 4: A Multinomial Naive Bayes Classification Approach for Customer Feedback Analysis task, 2017, Proceedings of the 8th International Joint Conference on Natural Language Processing, Shared Tasks, pp. 161–169.

[18] Tobias Cabanski, Julia Romberg, Stefan Conrad, HHU at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Data using Machine Learning Methods, 2017, Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval), pp 832–836.