ASEE 2022 ANNUAL CONFERENCE
Excellence Through Diversity
MINNEAPOLIS, MINNESOTA, JUNE 26TH-29TH, 2022 ASEE

Paper ID #37665

Using Sentiment Analysis to Evaluate First-year Engineering Students Teamwork Textual Feedback

Abdulrahman M Alsharif (Graduate Research Assistant)

Andrew Katz (Assistant Professor)

David B Knight (Associate Professor and Special Assistant to the Dean for Strategic Plan Implementation)

Saleh Zayed Alatwah (Data Scientist)

Using Sentiment Analysis to Evaluate First-year Engineering Students' Teamwork Textual Feedback

Abstract

Sentiment analysis (SA) is used in multiple disciplines to evaluate textual data and has become a popular topic in educational research, with a growing body of published work. SA has been employed in educational research to investigate student satisfaction, attitudes, topics of concern, or to evaluate instructors' teaching performance. However, there has been little discussion of applying SA as an assessment approach to evaluate teamwork textual feedback (i.e., students rate their teammates by writing comments on them) in engineering. The purpose of this research is to investigate the possibility of using SA as a method for evaluating collaborative textual feedback (e.g., comments) from students and to show its potential in assisting teachers in evaluating teamwork dynamics in their classrooms.

Teamwork is a key skill in engineering. With the rising complexity and magnitude of the challenges engineers handle, teamwork has become increasingly important. This is reflected in the Accreditation Board for Engineering and Technology accreditation student outcome criteria 3.5, which specifically highlights an ability to effectively function on teams. Engineering education literature further demonstrates the importance and the responsibility of faculty involvement in the development of effective teamwork. To assess teamwork functionality, instructors can distribute a survey among teams for team members to provide feedback about each other. This kind of feedback is helpful not only for that specific team and class but also for identifying broader, systematic trends in engineering student teams. Often the textual feedback is gathered at the end of a semester, and evaluating these responses can identify useful insights for improving teaching approaches. Unfortunately, in many cases, such surveys also can go underutilized. Large amounts of textual data often are not compatible with traditional analytic methodologies, but we claim that these huge amounts of textual data have the potential to deliver unique insights to educators and researchers.

We investigate SA as a potential method for analyzing a large corpus of student feedback responses about their team members and test this concept using a sample from 53,088 student responses from a first-year engineering course. The purpose of this research is to propose SA as an assessment approach to evaluate students' teamwork comments. According to research, potential issues that first-year engineering students frequently face when working in teams include teammates not performing their share of their work, tardy teammates, domineering teammates, and some team members being excluded from major tasks. SA has the potential of identifying team members' biases, and it can provide quick feedback and provide real-time insights into the teamwork environment (e.g., positive, neutral, or negative environment). Research shows that conflicts between team members are common, and constructive feedback is critical in the development of students both as individuals and teammates. Furthermore, insights from SA have the ability to augment instructor evaluations of teamwork. By analyzing survey comments based on the comment writer and the individual about whom the comment was made.

Introduction

Having emerged as an active study field in natural language processing (NLP) since its inception in the early 2000s, sentiment analysis (SA) has developed into one a widely used tool today in NLP (Liu, 2012). In the last few years, SA has been used in education to understand learning processes, student performance, course abandonment, teaching processes, and course satisfaction (Mite-Baidal et al., 2018). Sentiment analysis is an NLP approach that is often associated with opinion mining (Zhang & Liu, 2017). SA is a technique for determining the polarity of comments and is used in combination with opinion mining to extract information (e.g., the positive, neutral, and negative emotions in comments) based on an individual's expressed opinion (Balahadia et al., 2016). SA is mostly used in organizations as a system for gathering and evaluating client feedback, such as blogs, comments, reviews, or tweets, which are often employed in enterprises (G & Chandrasekaran, 2012). In industrial settings, this information can then help to improve the quality of a certain service or product for consumers. In academic settings, an application area could be with respect to the practice of gathering feedback from students about their team members via peer assessment, which is important since research on student teams suggests that peer assessment can promote effective teamwork (Bacon, 2005; Chowdhury & Murzi 2019; Riebe et al., 2016). These peer assessments often contain both rating and open-ended questions that reveal students' opinions about their interactions with their teammates.

In reality, many researchers utilize quantitative metrics acquired via surveys for this kind of data collection, and they depend on predefined rubric guidelines that may not consider certain aspects of students' opinions (Balahadia et al., 2016). It is important to note that instructors need to monitor not only team progress toward completion of tasks and activities, but also to evaluate students' self-assessment and peer-assessment on a frequent basis. If an instructor teaches multiple classes simultaneously and receives hundreds of self-and peer-assessments, it will be difficult for one person to read and evaluate all of them in a timely manner. However, sentiment analysis enables instructors to rapidly look into students' opinions on many elements of their collaboration (e.g., teammates, workload, and task complexity) and quickly identify potential issues without having to sift through hundreds of comments at once. We suggest that SA is a viable technique to quickly examine students' peer assessment and gather a high-level view of group dynamics.

In the present research paper, we explored SA as a proposed method for analyzing a large data set of student feedback responses (n=53,088) about their team members gathered throughout the semester in a first-year engineering course. Specifically, we used SA to categorize students' comments as positive, neutral, or negative. We examined the proportions of positive, neutral, and negative sentiments indicated in students' comments, broken down by demographic variables of both the student who made the comment and the student about whom the comment was made. With these comments, we then derive an overall positive or negative evaluation of the team member. The purpose of this paper is to discuss our general method to evaluate students' peer assessments on teamwork and to show the potential of using SA and insights into engineering student team dynamics gained from applying the method.

Teamwork

Due to the rising complexity and magnitude of the challenges engineers handle, the National Academy of Engineering report emphasizes the necessity of engineers to work in multi-disciplinary teams as a critical engineering ability (NAE, 2012). In the 21st-century, it is desirable and necessary for today's workforce to collaborate productively and successfully with teams (Riebe et al., 2016). The value of teamwork is also reflected in studies indicating that when organizations seek to hire new talent, future employers search for individuals who can work well in engineering teams (Chowdhury & Murzi, 2019; NAE, 2012; Passow, 2012). To underscore the importance of teamwork, one of the requirements for accreditation by the Accreditation Board for Engineering and Technology (ABET) is that students demonstrate the capacity to function in high-performing teams (ABET, 2017).

Importance of Peer-assessment in Teamwork

The literature on engineering education addressed a number of strategies to ensure effective teamwork in engineering where it focuses on improving student teaming skills and faculty responsibilities for establishing successful teamwork (Chowdhury & Murzi, 2019). Instructional strategies include self-assessment and peer-assessment that can be used to better understand the effectiveness of student teams (Hoffart, & O'Neill, 2016; Helmi et al., 2016). However, students' perceptions about team composition are dependent on distributive fairness of workload (or tasks) and what would provide the highest reward for their engagement in teamwork (Riebe et al., 2016). This can create phenomena such as social loafing (i.e., individuals expending less effort when working in a group than when working alone), which is students' main concern when they work in groups (Riebe et al., 2016). In classroom teamwork settings social loafing happens when one or more team members refuse to contribute their fair share to a team endeavor, benefiting from the work of others and producing resentment (McCorkle et al., 1999). To counter this problem, research suggests instructors should emphasize action processes and include a clear provision for individual accountability in assessing their teammates (Bacon, 2005; Riebe et al., 2016). Peer assessments are critical for instructors to employ action processes since they aid in not only understanding and analyzing teamwork dynamics, but also in reducing social loafing in group work.

Sentiment Analysis Approaches

In general, there are two main methods for SA - those that are based on either machine learning, or those based on lexicographical features corresponding to hard-coded rules (Gonçalves et al., 2013; Tao et al., 2019). Sentiment detection is often given as a binary option (e.g, positive or negative) between two outcomes when using the machine learning approach (Gonçalves et al., 2013). This approach requires either a supervised learning approach (e.g., training classifiers) (Pang et al., 2002) or an unsupervised learning approach to collect sentiments from texts (Tao & Fang, 2020). With these machine learning-based approaches, it is important to note the potential applicability to data outside the domain on which the models were trained. This can translate to limited application because of the lack of easily accessible labeled data; however, it could be beneficial when training models for specific purposes and situations (Gonçalves et al., 2013). The downside of this approach is that it is common for a classifier trained on labeled data in one domain to perform badly in another one (Liu, 2012). In educational research, historically, the most utilized approaches are the lexicon-based approach, supervised machine learning, and unsupervised machine learning (Mite-Baidal et al., 2018). Every approach shares a similar goal which is to provide a statement about the sentiment expressed in the textual data. Approaches might vary in their accuracy in placing a sentence in a positive, negative, or neutral category.

Application of Sentiment Analysis

Statement analysis is often utilized in enterprises or organizations to understand their customer feedback about a product or service. It is widely used in social media because a fundamental characteristic of social media is that it allows anybody from anywhere in the globe to freely express their ideas and opinions without revealing their genuine identity or fear of negative consequences (Liu, 2012). Previous studies used SA on social media platforms (e.g., Blogs, Facebook Twitter, Tumbler, and Instagram) to enhance judgments, suggestions, and services (Drus & Khalid, 2019). Organizations are increasingly relying on the information contained in various social media platforms to make decisions (Liu, 2012). Other studies have used SA in a product or movie review to better understand their customers and make the appropriate decisions to enhance their products or services, respectively (Gursoy et.al., 2017). In this project, we apply this approach to teamwork data from an engineering education context to answer the following questions.

Research Questions

- RQ1) What are the most prevalent sentiments when using SA on teamwork peer-evaluation data?
- RQ2) What are the proportions of the sentiment expressed in team feedback by raters' demographics (e.g., race, gender, international status)?

Methods

We utilized sentiment analysis, in particular Valence Aware Dictionary for Sentiment Reasoning (VADER), as an approach to evaluate engineering students' feedback on their teammates. VADER is an unsupervised machine learning algorithm (Hutto & Gilbert, 2014); compared to an average group of 20 human raters for sentiment intensity, the algorithm performed well with an r = 0.881 correlation coefficient compared to the human group's had an r = 0.888. The way VADER works is by summing positive, negative, and neutral scores, which are then adjusted between -1 and +1 (i.e., +1 is the most positive and -1 is the most negative). VADER scoring system follows a dictionary-based approach where words are given a particular score based on their nature and their usage (e.g., positive, negative, and neutral). In this study, students' comments were split into two parts because some comments are lengthy and increase the noise for the AI and lead to misunderstanding. The total comments before splitting were n=53,088 and after the split was n=100,915 comments. For the splitting producers, we used a spacy sentence segmenter which split the sentence automatically after a full stop. Initially, multiple dictionary-based models were tested; of those tested, VADER performed best in our data set as determined by comparing a sample of the sentiment classifications with human ratings of those same comments.

Data Collection

The data were collected from The Center for Advanced Technology in Engineering Education (CATME), a non-profit center based at Purdue University's School of Engineering Education. CATME is a secure system that includes web-based tools to allow educators to use best practices in managing student teams. According to CATME's official website, the system has been used by over 1,470,772 students and 17,000 instructors. In this study, we focused on the qualitative part of the peer evaluation survey, specifically raters' comments about their team members in two first-year engineering courses (ENGR131 and ENGR132) from a public land-grant research university. The total number of students pre-split and the after the cleaning stage is 53,088 who committed during the Spring of 2017, 2018, and 2019. All not applicable (NA), none, nothing, and characters symbols comments were removed in a pre-processing step. Following this screening, the remaining comments were split into two, three, and more than four parts (n= 100915) prior to the analysis stage, the splitting was necessary to ensure that positive comments were separate from negative comments. It also helps VADER provide a more accurate sentiment. Table 1 shows the proportions of demographic variables.

Table 1. Distribution of Comments by Student Gender (N = 100,915)

Rater Demographic	Count (n)	Target Demographic	Count (n)
Female	25,280 (25%)	Female	23,778 (24%)
Male	73,714 (73%)	Male	74,988 (74%)
Others (Gender)	1,921 (1.9%)	Others (Gender)	2,149 (21%)

Table 1.1 Distribution of Comments by Student International Status (N = 100,915)

Rater Demographic	Count (n)	Target Demographic	Count (n)
International Student (No)	83,729 (83%)	International Student (No)	82,588 (82%)
Others (International Status)	1,426 (1.4%)	Others (International Status)	1,713 (1.7%)
International Student (Yes)	15,760 (16%)	International Student (Yes)	16,614 (16%)

Table 1.2 Distribution of Comments by Student Race (N = 100,915)

Rater Demographic	Count (n)	Target Demographic	Count (n)
Asian	20,464 (20%)	Asian	20,819 (21%)
Black	1,760 (1.7%)	Black	2,140 (2.1%)
Declined	2,640 (2.6%)	Declined	2,513 (2.5%)
Hispanic	9,503 (9.4%)	Hispanic	9,699 (9.6%)
Native	97 (<0.1%)	Native	96 (<0.1%)
Others (Race)	4,768 (4.7%)	Others (Race)	4,825 (4.8%)
White	61,683 (61%)	White	60,823 (60%)

The study focuses on sentiment of the rater's comments (e.g., review of the teammaber) toward other team members. Therefore, to better understand the distribution of comments by raters international status table 1.2 shows the racial breakdown of the international status.

Table 1.3 Race Proportions of Raters International Status (N = 100,915)

Race	Yes	No
Asian	10,294 (13%)	10,114 (64%)
Black	1,337 (1.6%)	423 (3%)
Declined	1,785 (2%)	855 (5%)
Hispanic	7,775 (9%)	1,716 (11%)
Native	64 (0.08%)	33 (0.21%)
Other	2,812 (3%)	675 (4%)
White	59,662 (71%)	1,944 (13%)
Total	83,729	15,760

The total number of comments per- and post-split varies. We believe the difference is related to the data cleaning processes before and after splitting, as well as the fact that students' comments fluctuate in length, and Python's spacy sentence segmenter splits the sentence after the full stop, resulting in several splits per comment. Table 2 shows the distribution of comments per semester and table 3 shows descriptive statistics of students' comments.

Table 2 Distribution of Comments Per Semester

Semester	Pre-split	Post-split
Spring 2017	17,828 (34%)	34,584 (34%)
Spring 2018	16,562 (31%	34,967(35%)
Spring 2019	18,698 (35%)	31,364 (31%)
Total	53,088	100,095

Table 3. Descriptive statistics of students' comments

Comments	Min	Max	Mean	Median
Original	1	707	46	38
Split	1	97	15	14

Data Analysis and Results

In the data analysis stage, the data was evaluated by using Valence Aware Dictionary for Sentiment Reasoning (VADER) in Python 3.7. 6 to generate polarity for students' comments and R version 3 to visualize and analyze the data set. The polarity score can vary from -1 to 1, with any negative polarity indicating a negative statement and any positive polarity indicating a positive statement. To answer RQ1 we removed the stop words from all comments and provided both a bigram (e.g., most frequent two words used by students) for the positive sentiments and the top-50 words used by all students in all sentiments (e.g., positive, neutral, negative). For the negative sentiment, we decided to use the four words because two or three words cannot convey the same meaning as the full split sentence. To answer RQ2, we followed a descriptive approach by providing a proportion of sentiment across raters and target demographics (e.g., race, gender, international status). Our results showed that students complained about teammate performance, workload, and task difficulties. It also shows the potential of sentiment analysis in evaluating peer assessment surveys. Figure 1. shows a bar graph of the distribution of classes of student sentiment.

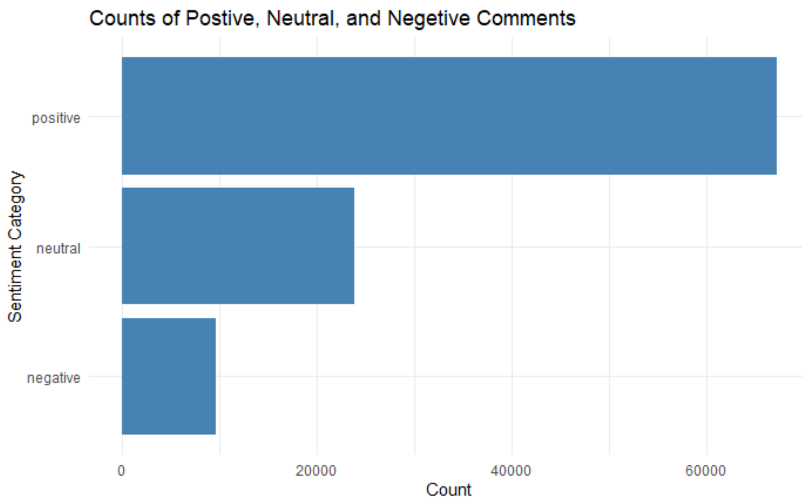


Figure 1. Proportions of Students' Sentiments

Figure 1 shows the counts of the negative, neutral and positive comments made by students. The majority of the comments are positive at over 60,000 comments; there were around 25,000 neutral comments and 10,000 negative comments. To further understand these sentiments and answer RQ1, we investigated the frequent words used in the overall data set and then did the same procedures to the negative and positive sentiments to visualize the most frequent words used. Figure 2 shows the top 50 frequent words used in the overall data set.

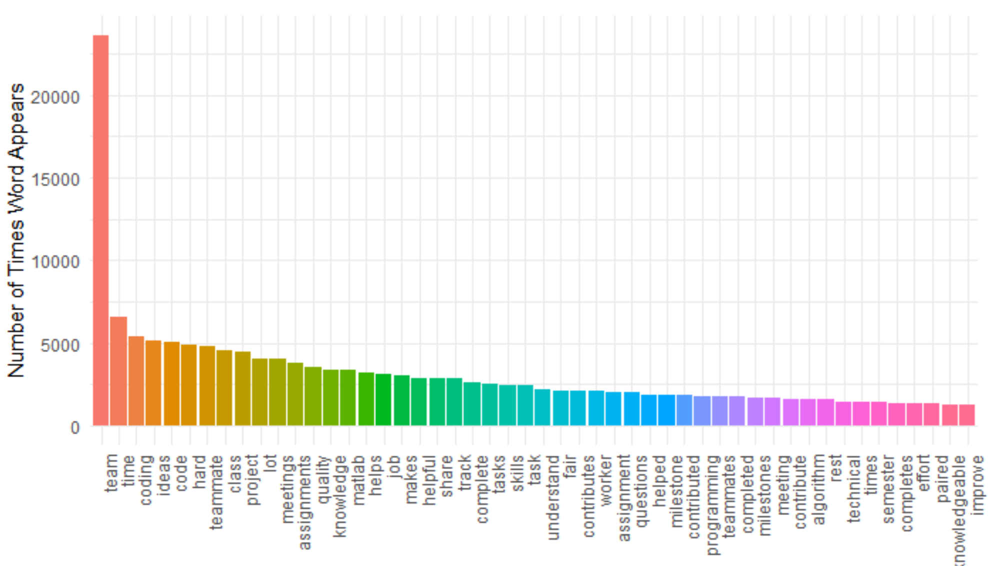


Figure 2. Raters' Top 50 words in the Overall Teamwork Peer assessment.

Compared to the other words, the word "team" showed up 28234 times in students' comments. The second most used word is "time", and the third most used word is "coding."

Furthermore, the word “time” showed up 7629 times and the word coding showed up 6160 times. There is an enormous gap between the word team and every other word used by students. It is important to understand the sentence that the world showed up in because the meaning and the delivery of the word “team” can vary from positive, negative, and neutral. For example, a positive comment from the data set is “a great team member”, a negative comment is “I don't like my team” and a neutral comment is “He didn't speak a lot in the team meetings.”

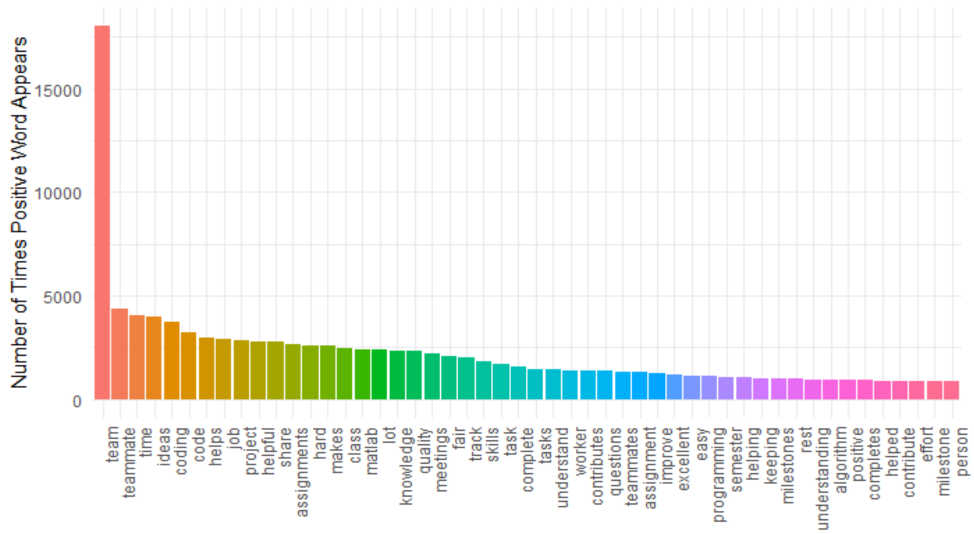


Figure 3. Raters’ Top 50 words in Teamwork Peer-assessment (Positive Sentiment)

The top 50 words in the positive sentiment are similar to the overall top 50 words in students' peer assessment. This is not surprising because in figure 1 the positive sentiment is the majority of comments in the overall data set. Students expressed that their teammates help in the work, theme work ethic (e.g., hard worker, excellent work, quality work, etc.), understanding of coding and programming, task difficulty (e.g, easy and not hard), and time management in task completion. The most used bigram associated with the positive sentiment is shown in table 4.

Table 4. Most frequent Bigram in the positive sentiment

Bigram	**Count (n)**
team_member	2787
makes_sure	2306
fair_share	2075
always_willing	1699
great_teammate	1622
good_teammate	1610

Table 4 Continues

work_done	1548
good_job	1525
make_sure	1482
quality_work	1463

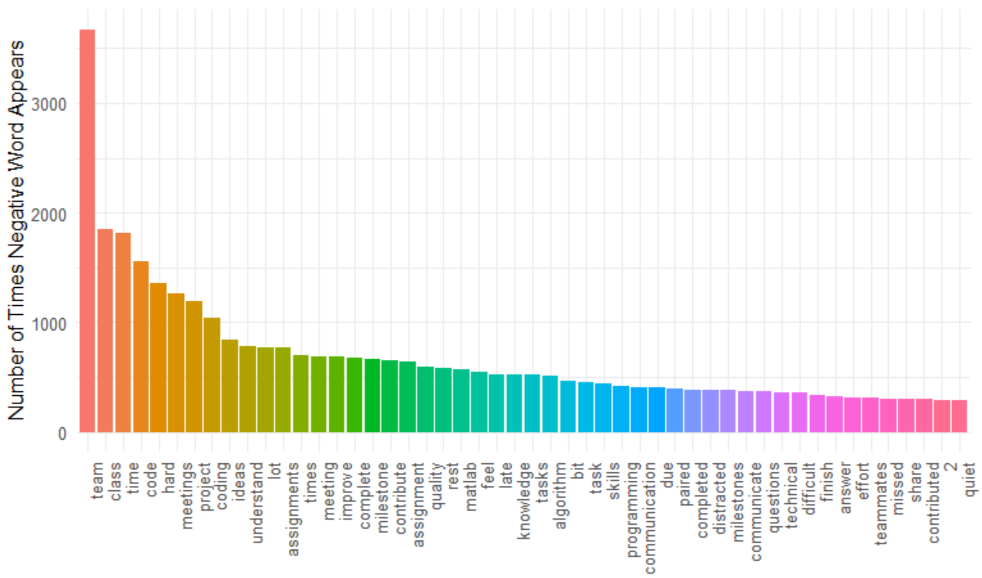


Figure 4. Raters' Top 50 words in Teamwork Peer-assessment (Negative Sentiment)

In Figure 4 the word "team" is also the most commonly used word in the Negative Sentiment. However, the sentence in which the word is used is negative. In the negative sentiment, students complained about the difficulty of the class, time in terms of finishing the project and other tasks, not understanding the code or programming, teammates missing meetings, teammates' lack of contribution, lack of communication between teammates, and late submissions. We decided not to use bigram or trigram because two or three words cannot express the same meaning as the full split sentence. Especially when it comes to negative words. For example, quality work shows up 101 times and from the negative sentiment data an example is "When she does work, it is not of a good quality and needs to be redone." Bigrams are shown in table 5.1. Instead, we decided to show examples of four words used by the students. The four words are crucial in assisting the teacher in gaining a better grasp of the team dynamics and the students' concerns. Table 5 shows the most frequent four-gram.

Table 5. Most frequent four-gram in the positive sentiment

Four-gram	Count (n)
can_sometimes_get_distracted	3
problem_set_paired_programming	3
sometimes_can_get_distracted	3
missed_several_team_meetings	4
problem_sets_problem_sets	3
problem_sets_assigned_contributed	3
can_get_distracted_talking	4
required_problem_sets_problem	3
set_paired_programming_questions	3
expects_high_quality_work	3

Table 5.1 Most frequent Bigram in the positive sentiment

Bigram	Count (n)
team_meetings	500
team_member	491
problem_sets	297
problem_sets	155
work_done	148
paired_programming	146
really_hard	126
quality_work	101

Table 5.1 shows bigrams from negative sentiment and the meaning of each bigram is unclear and vague. Therefore, we decided to use four-gram instead to somewhat understand the context of the negative sentiment.

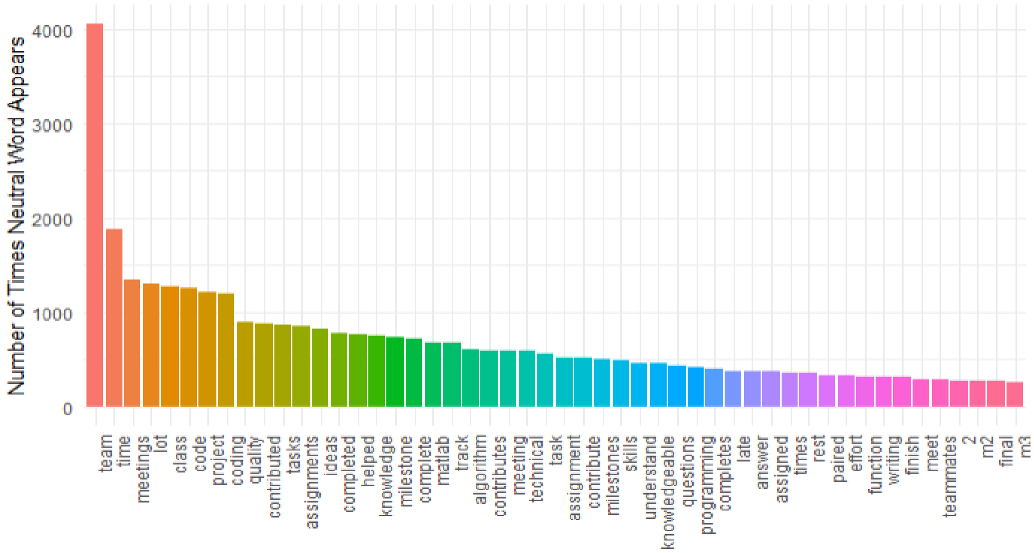


Figure 5. Raters' Top 50 Words in Teamwork Peer-assessment (Neutral Sentiment)

Figure 5. Illustrates raters' most used word across the neutral sentiment in teamwork peer assessment. It can be seen that the word team is still the most used word across all three sentiments. It appears that students use the word team the most when they review their teammates negatively, neutrally, or positively. Unlike figure 3 and figure 4, to understand the context of the figure 6, table 6 shows the most used bigrams across the neutral sentiment. Where the most used bigrams revolve around work and the bigram with the word team on it comes last. There are many similarities between figures 3, 4, and 5. The words time, code, and project show in the top 5 most frequent words used by students to access all sentiments. Furthermore, bigrams can be used as keywords to search for student comments. For example, "technical_brief" is the most used neutral sentiment by students (See Table 6). Examples of the students' split comments using "technical_brief" as a keyword. One student said, "She did the work expected of her, particularly with the technical brief, and always communicated if she wouldn't be at the meetings. Another student said, "Afterwards, he also read over the entire technical brief and made adjustments."

Table 6. Most Frequent Bigram in the Neutral Sentiment

Bigram	**Coun (n)**
technical_brief	443
work_done	394
quality_work	374
team_meetings	325
always_willing	277
team_member	200
team_members	168
answer_sheet	152

In order to visualize students' comments across rater races, gender, and international status, we utilized multiple bar graph plots to answer RQ2. Figure 6 shows the proportions of sentiment differences among target races.

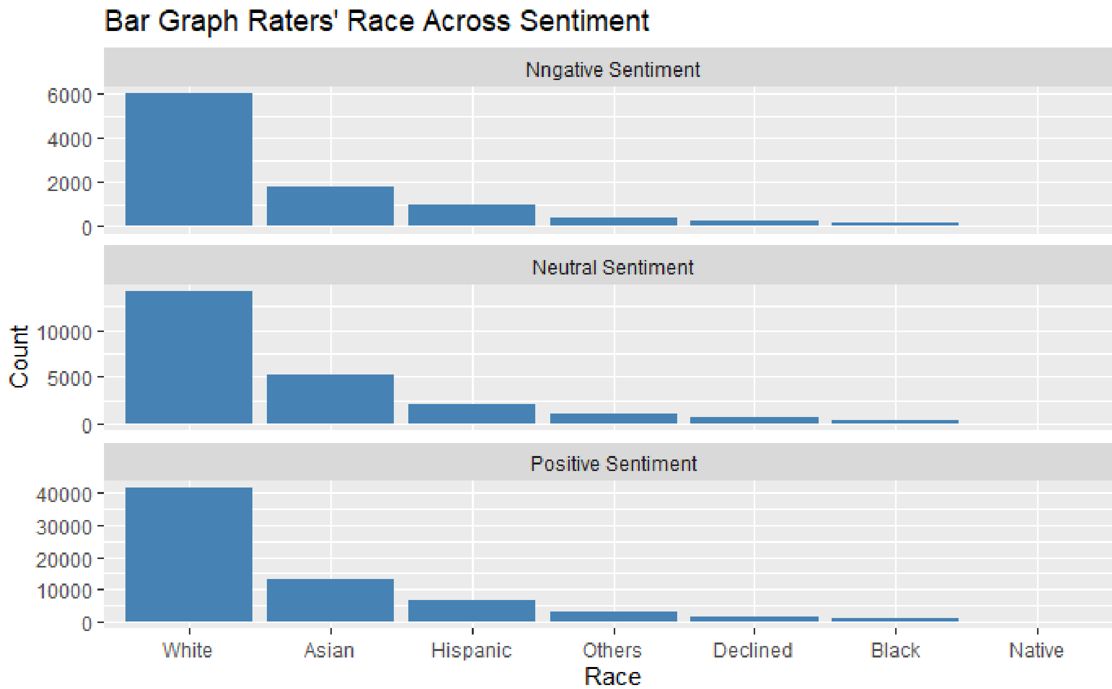


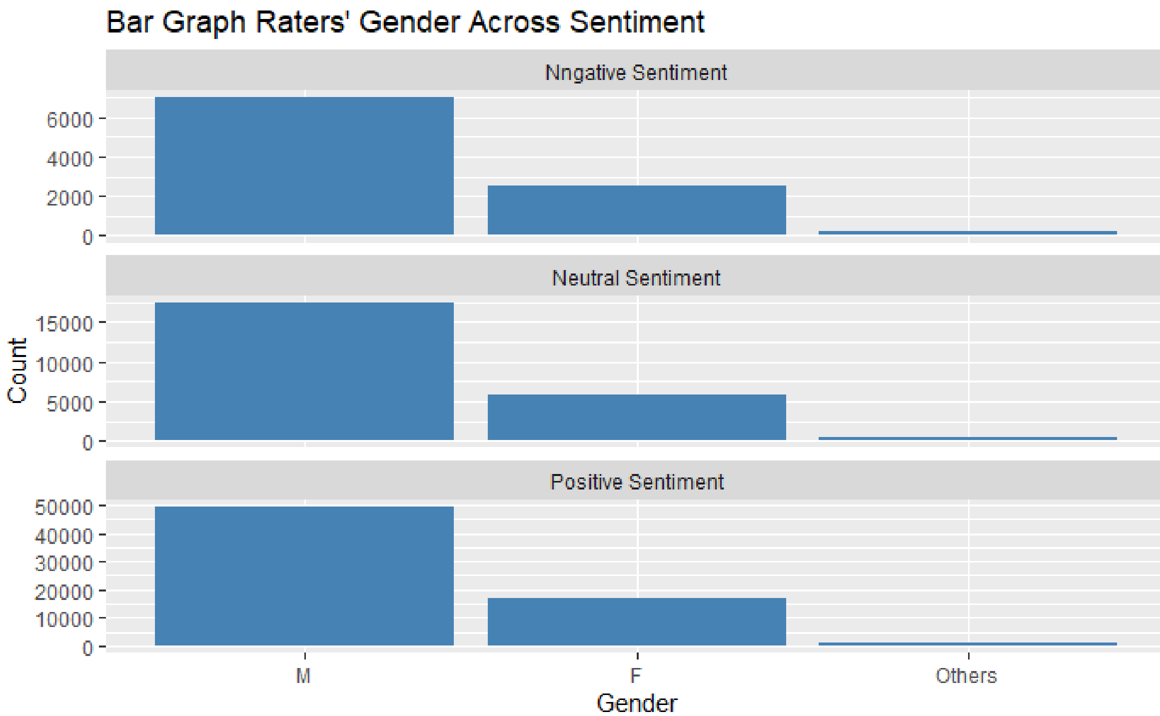
Figure 6. Bar Graph Raters' Race Across Sentiment

Figure 6. shows the proportions of sentiment used by raters based on their race. Most negative sentiments from raters are from White (40,000 positive comments) and Asians (15,000 positive comments). Most neutral and positive sentiments are also from White, and Asian raters. The results are not surprising because White counts for 61% of the data set and Asin only 20% see table 1. Where Hispanics come third with approximately 10,000 positive comments, 5000 neutral comments, and 2000 negative comments. For black, declined, native, and others races table 7 will shows breakdown of their proportions

Table 7 Proportions of Underrepresented Minorities Across Sentiments

Race	**Negative**	**Neutral**	**Positive**
Black	155 (0.15%)	453 (1.14%)	1152 (1.14%)
Declined	286 (0.66%)	666 (1.67%)	1688 (1.67%)
Native	8 (0.03%)	29 (0.06%)	60 (0.06%)
Others	432 (1.11%)	1120 (1.11%)	3216 (3.19%)

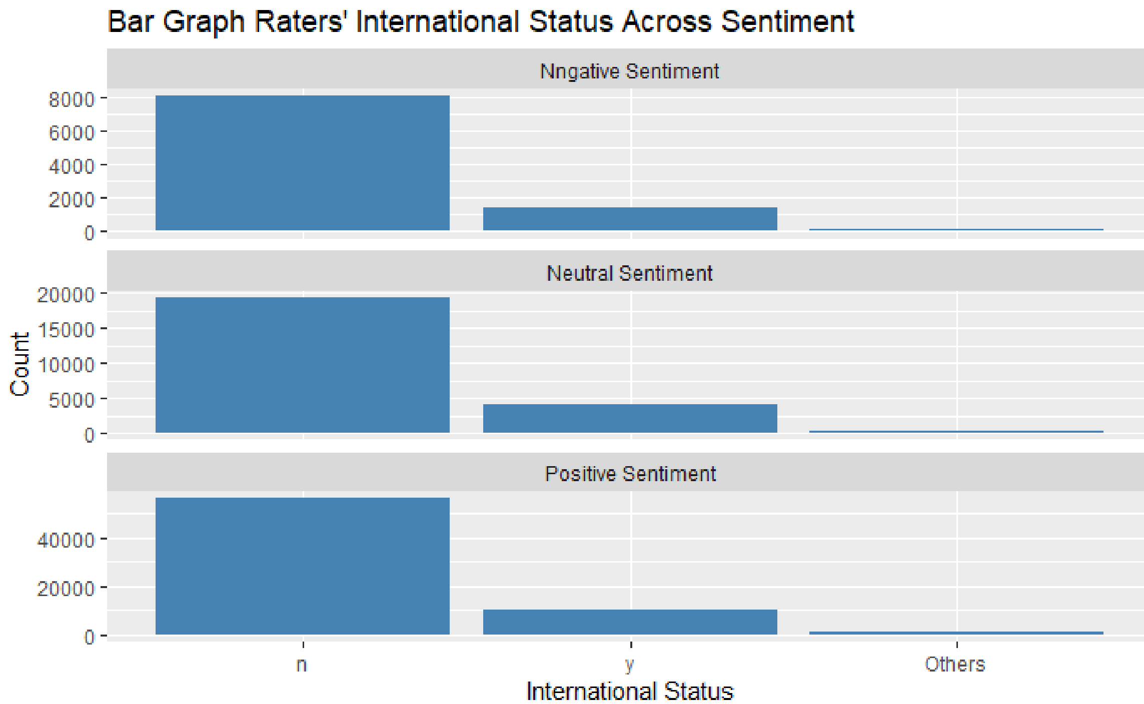
*Proportions /100,915 (total number of comments)



*M= Male, F= Female

Figure 7. Bar Graph Raters' Gender Across Sentiment

Figure 7. represents the proportions of sentiment used by raters' gender. The most frequent positive, neutral and negative sentiments are by Male raters while female raters come second with the same sentiment order and other genders come last. In the three gender categories, the positive comments from the rater count for the most sentiment used by raters. Male have more negative comments compared to females and other genders. It should be noted that male counts for the majority of the data.



*n=No, y= Yes

Figure 8. Bar Graph Raters' International Status Across Sentiments

Figure 8 displays the sentiment proportions of raters' international status. The vast majority of raters are not international students, accounting for 83% of the data (see table 1). Positive, neutral, and negative raters' sentiments are presented in descending order. Students with international status come second with similar sentiment rankings, with positive comments accounting for approximately 10000, neutral comments accounting for approximately 4000, and negative comments accounting for approximately 2000.

Discussion

Prior sentiment analysis research covered understanding learning processes, student performance, course abandonment, teaching procedures, and course satisfaction in education (Mite-Baidal et al., 2018). There is little to no research on the use of sentiment analysis in teamwork settings to evaluate peer-assessment surveys. Assessing teammates via a survey instrument is comparable to evaluating their peers' work quality. We evaluated sentiment analysis as a tool for evaluating engineering student teamwork peer assessment in this paper, in particular for the survey's open-ended section. The approach taken in this article is called Valence Aware Dictionary for Sentiment Reasoning (VADER), which is a combination of a sentiment lexicon approach and an NLP algorithm that is dictionary sensitive. During the analysis phase, the dictionary encountered some ambiguity with certain terms (e.g., hard and tough) and assigned these terms a negative polarity score, even when the connotation may have been positive (e.g., hard worker and tough worker). We thus had to manually program the algorithm to dedicate bigrams that are similar to hard-working or tough workers and give them a positive polarity. In terms of sentiment across comments made by underrepresented minorities, they are similar to those made by white and Asian. It is very difficult to distinguish the difference due to the size of the data. A smaller data set will be more manageable and easier to interpret. The purpose of this study is to purpose SA as an evaluation method and to show its potential. As a result, comments polarity should be used on a smaller dataset so that gender and race differences can be determined more accurately.

Table 1 lists the most frequently used positive bigram by students that are associated with positive teamwork characteristics. For instance, the bigram "fair share" occurred 2075 times, while "quality work" occurred 1463 times. However, in the negative trigram, terms such as "team_member" are meaningless due to the comment's ambiguous context. We used four grams to convey a negative sentiment, and the frequently used phrase "begin to make sense" (e.g., can sometimes get distracted). Despite the fact that we have a sizable data set, it is difficult for students to provide the exact same four-word sentence as comments for their peers. To answer research RQ2, we used different bar graphs to display the proportion of sentiment across raters' demographics (e.g., race, gender, international status). The majority of all sentiment (positive, negative, and neutral) is attributed to White, Asian, and male respondents, as they represent the majority of the data set. In a typical classroom setting with a manageable data set (i.e., average classroom size), these results may appear significantly different and may provide critical information to the instructor in a timely manner when comparing across multiple variables is challenging for a human to process (e.g., textual responses by demographics).

Conclusion

The objective of this study is to use sentiment analysis as a method to swiftly analyze teamwork peer assessments in a first-year engineering course. Our data set contains 53,088 full-length comments; after dividing the comments, we had 100,915 total comments. Vader, the sentiment analysis technique we employed in this study, as well as the others, are not particularly accurate when it comes to long passage sentiment. As a result, we've divided the comments so that Vader can evaluate them accurately. The assessment's open-ended question offers vital information on team member performance and behavior and comprehending this information will provide the teacher with the ability to provide enough support to students, assuring improved team dynamics and working environment.

We provide descriptive statistics about the proportions of sentiments conveyed in students' comments (e.g., positive, neutral, or negative comments) split by demographic characteristics (e.g., gender, race, and international status) of both the student making the comment and the student about whom the comment was made. Our findings suggest that sentiment analysis can assist instructors in comprehending collaborative dynamics more quickly, providing them with a better sense of how relationships are developing in their classroom teams. SA can also be used to identify positive, neutral, and negative words used by students, allowing the teacher to take measures based on the feedback and assist students to work in a more productive learning environment.

According to our findings, students with a positive attitude rate their team members positively. Reviewing the positive bigrma we can result in students demonstrating good work quality, topic knowledge, understanding of task complexity, and time management in performing tasks. We opted not to utilize bigram or trigram for the negative attitude since the top words are readily misconstrued. For example, the word hard-working or quality work is used in a negative way rather than a positive context in the full split sentence, so we decided to assess them by looking at the four words used in the comments and skimming through them to understand the context of the top 50 words in the negative sentiment. We also demonstrated the potential of using this kind of approach to analyzing textual data is the ability to cross-analyze large amounts of data with other variables, such as race, gender, and international status as demonstrated here.

Future Work

For future work, we suggest that SA be used on a smaller data set to spot differences between variables by investigating polarity using statistical tests such as ANOVA or Chi-square across multiple demographic variables, which will provide valid and accurate results. We also recommend evaluating other teamwork settings in engineering classrooms (e.g., design courses) and could be used to explore long textual feedback (e.g., journal reflections).

Limitations

There are a few limitations to be mentioned, sentiment analysis can be helpful in accelerating the speed of evaluation and can provide the instructor with a glimpse of what is happening unlike qualitative coding by humans which is more meritable and accurate. The approach of sentiment analysis can vary based on the type of data and cannot provide an accurate sentiment on very long passages. The data set we used was dominated by white ethnicity and Male gender. The collected data had a lack of gender options. Finally, in the raw data, the frequency of students' peer assessment evaluations is known.

References

ABET. (2017). 2018-2019 Criteria for Accrediting Engineering Programs. Retrieved from https://www.abet.org/wp-content/uploads/2018/02/E001-18-19-EAC-Criteria-11-29-17.pdf

Balahadia, F. F., Fernando, M. C. G., & Juanatas, I. C. (2016). Teacher's performance evaluation tool using opinion mining with sentiment analysis. *2016 IEEE Region 10 Symposium (TENSYMP)*. https://doi.org/10.1109/TENCONSPRING.2016.7519384

Chowdhury, T., & Murzi, H. (2019, July 10). *Literature Review: Exploring Teamwork in Engineering Education*.

Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013). Comparing and combining sentiment analysis methods. *Proceedings of the First ACM Conference on Online Social Networks - COSN '13*, 27–38. https://doi.org/10.1145/2512938.2512951

Hutto, C. J., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. 10.

Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

McCorkle, D., Reardon, J., Alexander, J., Kling, N., Harris, R., Iyer, R. V. (1999). Undergraduate marketing students, group projects, and teamwork: The good, the bad and the ugly? Journal of Marketing Education, 21, 106-117

Mite-Baidal, K., Delgado-Vera, C., Solís-Avilés, E., Espinoza, A. H., Ortiz-Zambrano, J., & Varela-Tapia, E. (2018). Sentiment Analysis in Education Domain: A Systematic Literature Review. In R. Valencia-García, G. Alcaraz-Mármol, J. Del Cioppo-Morstadt, N. Vera-Lucio, & M. Bucaram-Leverone (Eds.), *Technologies and Innovation* (pp. 285–297). Springer International Publishing. https://doi.org/10.1007/978-3-030-00940-3_21

National Academy of Engineering (NAE) 2012. Infusing Real World Experiences into

Engineering Education. Washington, DC: The National Academies Press. https://doi.org/10.17226/18184.

Riebe, L., Girardi, A., & Whitsed, C. (2016). A systematic literature review of teamwork pedagogy in higher education. Small Group Research, 47(6), 619–664.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 79–86. https://doi.org/10.3115/1118693.1118704

Passow, H. J. (2012). Which ABET Competencies Do Engineering Graduates Find Most Important in their Work? Journal of Engineering Education, 101(1), 95–118. https://doi.org/10.1002/j.2168-9830.2012.tb00043.x

Gursoy U. T., Bulut, D., and Yigit, C. (2017). “Social Media Mining and Sentiment Analysis for Brand Management.” Global Journal of Emerging Trends in e-Business, Marketing and Consumer Psychology 3 (1): 497-551

Tao, J., & Fang, X. (2020). Toward multi-label sentiment analysis: A transfer learning based approach. *Journal of Big Data*, *7*(1), 1. https://doi.org/10.1186/s40537-019-0278-0

Tao, J., Zhou, L., & Feeney, C. (2019, January 8). *I Understand What You Are Saying: Leveraging Deep Learning Techniques for Aspect Based Sentiment Analysis*. https://doi.org/10.24251/HICSS.2019.057

Zhang, L., & Liu, B. (2017). *Sentiment Analysis and Opinion Mining*. WorldCat.org. https://doi.org/10.1007/978-1-4899-7687-1_907