# SOCIAL LISTENING USING PREDICTIVE ANALYTICS

[1]S.Leema Mary and [2]S.Benita Deepthi
[1]Associate Professor and HOD Mathematics, [2]ME Computer Science and Engineering
[1]Department of Mathematics,
[1]Government College of Engineering, Salem, India.

***Abstract:*** Students' informal conversations on social media (e.g. Twitter, Facebook) shed light into their educational experiences opinions, feelings, and concerns about the learning process. Data from such an instrumented environments can provide valuable knowledge to inform student learning. Analyzing such data, however, can be challenging. The complexity of students' experiences reflected from social media content requires human interpretation. However, the growing scale of data demands automatic data analysis techniques. In this paper, we developed a workflow to integrate both qualitative analysis and large-scale data mining techniques. We focused on engineering students' Twitter posts to understand issues and problems in their educational experiences. We first conducted a qualitative analysis on samples taken from about 25,000 tweets related to engineering students' college life. We found engineering students encounter problems such as heavy study load, lack of social engagement, and sleep deprivation. Based on these results, we implemented a multi-label classification algorithm to classify tweets reflecting students' problems. We then used the algorithm to train a detector of student problems from about 35,000 tweets streamed at the geo-location of Purdue University. This work, forth first time, presents a methodology and results that show how informal social media data can provide insights into students' experiences.

***Index Terms*****:** Social Listening, Predictive Analysis, Automatic algorithms, Na€ıve Bayes

## 1. INTRODUCTION

SOCIAL media sites such as Twitter, Facebook, and You-Tube provide great venues for students to share their experiences, vent emotion and stress, and seek social support. On various social media sites, students discuss and share their everyday encounters in an informal and casual manner. Students' digital footprints provide vast amount of implicit knowledge and a whole new perspective for educational researchers and practitioners to understand students' experiences outside the controlled classroom environment. This understanding can inform institutional decision-making on interventions for at-risk students, improvement of education quality, and thus enhance student recruitment, retention, and success [1]. The abundance of social media data provides opportunities to understand students' experiences, but also raises methodological difficulties in making sense of social media data for educational purposes. Just imagine the sheer data volumes, the diversity of Internet slang, the unpredictability of location and timing of students posting on the web, as well as the complexity of students' experiences. Pure manual analysis cannot deal with the ever-growing scale of data, while pure automatic algorithms usually cannot capture in-depth meaning within the data [2].

Traditionally, educational researchers have been using methods such as surveys, interviews, focus groups, and Classroom activities to collect data related to students' learning experiences [3], [4]. These methods are usually very time-consuming, thus cannot be duplicated or repeated with high frequency. The scale of such studies is also usually limited. In addition, when prompted about their experiences, students need to reflect on what they were thinking and doing sometime in the past, which may have become obscured over time. The emerging fields of learning analytics and educational data mining (EDM) have focused on analyzing structured data obtained from course management systems (CMS), classroom technology usage, or controlled online learning environments to inform educational decision-making [1], [5], [6], [7]. However, to the best of our knowledge, there is no research found to directly mine and analyze student posted content from uncontrolled spaces on the social web with the clear goal of understanding students' learning experiences.

The research goals of this study are 1) to demonstrate a workflow of social media data sense-making for educational purposes, integrating both qualitative analysis and large scale data mining techniques as illustrated in Fig. 1; and 2) to explore engineering students' informal conversations on Twitter, in order to understand issues and problems students encounter in their learning experiences. We chose to focus on engineering students' posts on Twitter about problems in their educational experiences mainly because:

1. Engineering schools and departments have long been struggling with student recruitment and retention issues [8]. Engineering graduates constitute a significant part of the nation's future workforce and have a direct impact on the nation's economic growth and global competency [9].
2. Based on understanding of issues and problems in students' life, policymakers and educators can make more informed decisions on proper interventions and services that can help students overcome barriers in learning algorithms. The width of gray arrows represents data volumes—-wider indicates more data volume. Black arrows represent data analysis, computation, and results flow. The dashed arrows represent the parts that do not concern the central work of this paper. This workflow can be an iterative cycle.

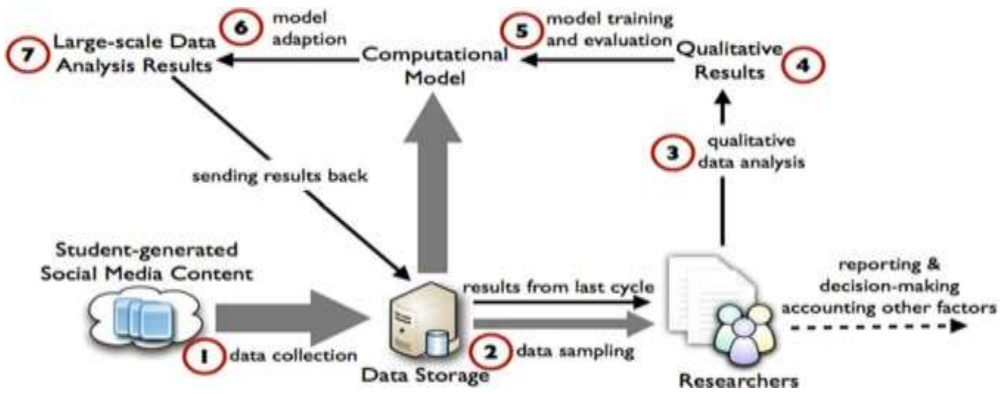



Fig. 1. The workflow we developed for making sense of social media data integrates qualitative analysis and data mining

3. Twitter is a popular social media site. Its content is mostly public and very concise (no more than 140 characters per tweet). Twitter provides free APIs that can be used to stream data. Therefore, we chose to start from analyzing students' posts on Twitter.

In this paper, we went through an exploratory process to locate the relevant data and Twitter hashtags (a Twitter hashtag is a word beginning with a # sign, used to emphasize or tag a topic). We collected 25,284 tweets using the hashtag #engineering Problems over a period of 14 months, and a second data set of 39,095 tweets using the geo-code (longitude and latitude) of Purge University, West Lafayette. This corresponds to step 1 in Fig. 1. Three researchers conducted an inductive content analysis on samples of the #engineering Problems data set, which corresponds to steps 2 and 3 in Fig. 1. In step 4, we found that major problems engineering students encounter in their learning experiences fall into several prominent categories. Based on these categories, we implemented a multi label Na€ıve Bayes classification algorithm. We evaluated the performance of the classifier by comparing it with other state-of-the-art multi-label classifiers (step 5). We used the classification algorithm to train a detector that could assist in the detection of engineering students' problems at Purdue University (step 6). The results (step 7) could help educators identify at-risk students and make decisions on proper interventions to retain them.

This paper makes two major contributions. First, it proposes a workflow to bridge and integrate a qualitative research methodology and large-scale data mining techniques. We base our data-mining algorithm on qualitative insights resulting from human interpretation, so that we can gain deeper understanding of the data. We apply the algorithm to another large-scale and unexplored data set, so that the manual method is augmented. We can keep refining the model based on further human feedback like the cycle illustrated in Fig. 1. Second, the paper provides deep insights into engineering students' educational experiences as reflected in informal, uncontrolled Environments. Many issues and problems such as study life balance, lack of sleep, lack of social engagement, and lack of diversity clearly emerge. These could bring awareness to educational pedagogy, policy-making, and educational practice.

The remainder of this paper is organized as follows: the next section reviews theory of public discourse online, related work on text classification techniques used for analyzing tweets, and data-driven approaches in education. Section 3 describes the data collection process (step 1 in Fig. 1). Section 4 details the inductive content analysis procedures and categories identified (steps 2, 3, and 4). Section 5 details the implementation of the Na€ıve Bayes multi-label classifier and the evaluation results (step 5). In Section 6, we show the comparison results of the Na€ıve Bayes classifier with the popular classifier— Support Vector Machine (SVM) and one of its variations Max-Margin Multi-Label (M3L) classifier. This is an additional evaluation of the classifier in step 5. In Section 7 we apply the Na€ıve Bayes classifier to the Purdue data set in order to demonstrate its application in detecting students' problems at a specific university (steps 6 and 7). Section 8 discusses limitations and possible future work, and Section 9 concludes this study.

## 2 RELATED WORK

### 2.1 Public Discourse on the Web

The theoretical foundation for the value of informal data on the web can be drawn from Goffman's theory of social performance [10]. Although developed to explain face-to-face interactions, Goffman's theory of social performance is widely used to explain mediated interactions on the web today [11]. One of the most fundamental aspects of this theory is the notion of front-stage and back-stage of people's social performances. Compared with the front-stage, the relaxing atmosphere of backstage usually encourages more spontaneous actions. Whether a social setting is front-stage or back-stage is a relative matter. For students, compared with formal classroom settings, social media is a relative informal and relaxing back-stage. When students post content on social media sites, they usually post what they think and feel at that moment. In this sense, the data collected from online conversations may be more authentic and unfiltered than responses to formal research prompts. These conversations act as a zeitgeist for students' experiences.

Many studies show that social media users may purposefully manage their online identity to “look better” than in real life [12], [13]. Other studies show that there is a lack of awareness about managing online identity among college students [14], and that young people usually regard social media as their personal space to hang out with peers outside the sight of parents and teachers [15]. Students’ online conversations reveal aspects of their experiences that are not easily seen in formal classroom settings, thus are usually not documented in educational literature. The abundance of social media data provides opportunities but also presents methodological difficulties for analyzing large-scale informal textual data. The next section reviews popular methods used for analyzing Twitter data.

### 2.2Mining Twitter Information

Researchers from diverse fields have analyzed Twitter content to generate specific knowledge for their respective subject domains. For example, Gaffney [16] analyzes tweets with hashtag #iranElection using histograms, user networks, and frequencies of top keywords to quantify online activism. Similar studies have been conducted in other fields including healthcare [17], marketing [18], athletics [19], just to name a few. Analysis methods used in these studies usually include qualitative content analysis, linguistic analysis, network analysis, and some simplistic methods such as word clouds and histograms. In our study, we built a classification model based on inductive content analysis. This model was then applied and validated on a brand new data set. Therefore, we emphasize not only the insights gained from one data set, but also the application of the classification algorithm to other data sets for detecting student problems. The human effort is thus augmented with large-scale data analysis. Below we briefly review studies on Twitter from the fields of data mining, machine learning, and natural language processing. These studies usually have more emphasis on statistical models and algorithms. They cover a wide range of topics including information propagation and diffusion [20], [21], [22], popularity prediction [23], [24], event detection [25], [26], topic discovery [27], [28], and tweet classification [29], [30], [31], [32], to name a few. Amongst these topics, tweet classification is most relevant to our study. Popular classification algorithms include Na€ıve Bayes, Decision Tree, Logistic Regression, Maximum Entropy, Boosting, and Support Vector Machine (SVM). Based on the number of classes involved in the classification algorithms, there are binary classification and multi-class classification approaches. In binary classification, there are only two classes, while multi-class classification involves more than two classes. Both binary classification and multi-class classification are single-label classification systems. Single-label classification means each data point can only fall into one class where all classes are mutually exclusive. Multi-label classification, however, allows each data point to fall into several classes at the same time.

Most existing studies on tweet classification are either binary classification on relevant and irrelevant content [31], or multi-class classification on generic classes such as news, events, opinions, deals, and private messages [32]. Sentiment analysis is another very popular three-class classification on positive, negative, or neutral emotions/opinions [33]. Sentiment analysis is very useful for mining customer opinions on products or companies through their reviews or online posts. It finds wide adoption in marketing and customer relationship management (CRM). Many methods have been developed to mine sentiment from texts. For example, both Davidov et al. [29] and Bhayani et al. [30] use emoticons as indicators to provide noisy labels to the tweets thus minimizing human effort needed for labeling. However, in the case of this paper, only knowing the sentiment of student-posted tweets does not provide much actionable knowledge on relevant interventions and services for students. Our purpose is to achieve deeper and finer understanding of students’ experiences especially their learning-related issues and problems. Even for a human judge to determine what student problems a tweet indicates is a more complicated task than to determine the sentiment of a tweet. Therefore, our study requires a qualitative analysis, and is impossible to do in a fully unsupervised way. Sentiment analysis is, therefore, not applicable to our study.

In our study, we implemented a multi-label classification model where we allowed one tweet to fall into multiple categories at the same time. Our classification was also at a finer granularity compared with other generic classifications. Our work extends the scope of data-driven approaches in education such as learning analytics and educational data mining.

### 2.3 Learning Analytics and Educational Data Mining

Learning analytics and educational data mining are data driven approaches emerging in education. These approaches analyze data generated in educational settings to understand students and their learning environments in order to inform institutional decision-making [34]. The present paper extends the scope of these approaches in the following two aspects. First, data analyzed using these approaches typically are structured data including administrative data (e.g., high school GPA and SAT scores), and student activity and performance data from course management systems (CMS) or virtual learning environments (VLE) such as Blackboard (http://www. blackboard.com/). For example, researchers at Purdue University created a system named Signals that mines student performance data such as time spent reading course materials, time spent engaging in course discussion forums, and quiz grades [35]. Signals give students red, yellow, or green alerts on their progress in the course in order to promote self-awareness in learning. Our study extends the data scope of these data-driven approaches to include informal social media data. Second, most studies in learning analytics and EDM focus on students’ academic performance [36], [37]. We extend the understanding of students’ experiences to the social and emotional aspects based on their informal online conversations. These are important components of the learning experiences that are much less emphasized and understood compared with academic performance from the Blackboard course management system.

## 3 DATA COLLECTION

It is challenging to collect social media data related to students’ experiences because of the irregularity and diversity of the language used. We searched data using an educational account on a commercial social media monitoring tool named Radian6 (http://www.salesforce. com/). The Twitter APIs [38] can also be configured to accomplish this task, which we later used to obtain the second data set. The search process was exploratory. We started by searching based on different Boolean combinations of possible keywords such as engineer, students, campus, class, homework, professor, and lab. We then expanded and refined the keyword set and the combining Boolean logic iteratively. The Boolean search logic grew very complicated eventually, but the data set still contained about 35 percent noise (during the month of November 2011, we retrieved 179 tweets, in which 63 were irrelevant to college students). Also, given that the data set was so small, we seemed to have ruled out many other relevant tweets together with the spam and irrelevant tweets.

From the limited number of relevant tweets we retrieved, we found a Twitter hashtag #engineering Problems occurring most frequently. Students used the hashtag #engineering Problems to post about their experiences of being engineering majors. This was the most popular hashtag specific to engineering students' college life based on the data retrieved using the Boolean terms. Using Radian6, we streamed tweets containing this hashtag for about 14 months (421 days) from November 1st, 2011 to December 25th, 2012. In total, we collected 25,284 tweets with the hashtag #engineering Problems posted from 10,239 unique Twitter accounts. Counting re-tweets, replies, and mentions, a total of 12,434 unique user accounts were involved. After removing duplicates caused by re-tweeting, there were 19,799 unique tweets in this data set. We also identified several other much less popular but relevant hashtags such as #lady Engineer, #engineering Majors, #switching Majors, #college Problems, and #nerd status. As a side note for future work, these hashtags can also be used to retrieve data relevant to college students' experiences. To demonstrate the application of the classification algorithm, we obtained another new data set using the geocode of Purdue University West Lafayette (40.428317, –86.914535) with a radius of 1.3 miles to cover the entire campus. From February 5th to April 17th, 2013, we obtained 39,095 tweets using the Twitter search API [38]. These tweets came from 5,592 unique user accounts. There were 35,598 unique tweets after removing duplicates. One reason we chose Purdue University as an example is that it is a large public university with a strong engineering student base. Over 27 percent (10,533/39,000) of the students at the West Lafayette campus are enrolled in the College of Engineering during the 2012-2013 academic year [39]. Nevertheless, the general approach we used can be applied to any institution and students in any major.

## 4 INDUCTIVE CONTENT ANALYSIS

Because social media content like tweets contain a large amount of informal language, sarcasm, acronyms, and misspellings, meaning is often ambiguous and subject to human interpretation. Rost et al. [2] argue that in large-scale social media data analysis, faulty assumptions are likely to arise if automatic algorithms are used without taking a qualitative look at the data. We concur with this argument, as we found no appropriate unsupervised algorithms could reveal in depth meanings in our data. For example, Latent Dirichlet Allocation (LDA) is a popular topic modeling algorithm that can detect general topics from very large scale data [40], [41]. LDA has only produced meaningless word groups from our data with a lot of overlapping words across different topics.

There were no pre-defined categories of the data, so we needed to explore what students were saying in the tweets. Thus, we first conducted an inductive content analysis on the #engineering Problems data set. Inductive content analysis is one popular qualitative research method for manually analyzing text content. Three researchers collaborated on the content analysis process.

### 4.1 Development of Categories

The lens we used in conducting the inductive content analysis was to identify what were the major worries, concerns, and issues that engineering students encountered in their study and life. Researcher A read a random sample of 2,000 tweets from the 19,799 unique #engineering Problems tweets, and developed 13 initial categories including: curriculum problems, heavy study load, study difficulties, imbalanced life, future and carrier worries, lack of gender diversity, sleep problems, stress, lack of motivation, physical health problems, nerdy culture, identity crisis, and others. These were developed to identify as many issues as possible, without accounting for their relative significances. Researcher A wrote detailed descriptions and gave examples for each category and sent the codebook and the 2,000tweet sample to researchers B and C for review. Then, the three researchers discussed and collapsed the initial categories into five prominent themes, because they were themes with relatively large number of tweets. The five prominent themes are: heavy study load, lack of social engagement, negative emotion, sleep problems, and diversity issues. Each theme reflects one issue or problem that engineering students encounter in their learning experiences.

We found that many tweets could belong to more than one category. For example, "This could very well turn into an all nighter...fyou lab report #no sleep" falls into heavy study load, lack of sleep, and negative emotion at the same time. "Why am I not in business school?? Hate being in engineering school. Too much stuff. Way too complicated. No fun" falls into heavy study load, and negative emotion at the same time. So one tweet can be labeled with multiple categories. This is a multi-label classification as opposed to a single-label classification where each tweet can only be labeled with one category. The categories one tweet belongs to are called this tweet's labels or label set.

### 4.2 Inter-Rater Agreement

Statistical measures such as Cohen's Kappa, Scott's Pi, Fleiss Kappa, and Krippendorf's Alpha are widely used to repor [47]. Thus students might benefit from more time to engage in social activities. Lack of social engagement is also intertwined with the stereotypical nerdy and anti-social image of engineers. Some students embrace the anti-social image (e.g., "well I suck at social interactions so this engineering will be good for me", and "People tell me I have to get a job and work with people. Why can't I just sit in my corner and do my math?"), while most others desire more social life as the examples above show. Loshbaugh and Claar [48] find that many students feel the nerdy and geeky culture of engineering is unwelcoming, which is detrimental for student recruitment and retention. Eventually, the society needs engineers who can work with people and solve problems that can benefit humanity [9], [49].

### *4.3* Negative Emotion

There are a lot of negative emotions flowing in the tweets. Admittedly, the hashtag #engineering Problems has a negative connotation. We only categorize a tweet as "negative emotion" when it specifically expresses negative emotions such as hatred, anger, stress, sickness, depression, disappointment, and despair. Students are mostly stressed with schoolwork. For example, "is it bad that before i started studying for my tests today that i considered throwing myself in front of a moving car??", "looking at my grades online makes me sick", "40 hours in the library in the past 3 days. I hate finals", and "I feel myself dying, #nervous". It is necessary for students to get help with how to manage stress and get emotional support.

#### 4.3.1 Sleep Problems

Our analyses find that sleep problems are widely common among engineering students. Students frequently suffer from lack of sleep and nightmares due to heavy study load and stress. For example, "Napping in the common room because I know I won't sleep for the next three days", "If I don't schedule in sleep time, it doesn't happen", and "I wake up from a nightmare where I didn't finish my physics lab on time". Chronic lack of sleep or low-quality sleep can result in many psychological and physical health problems; therefore this issue needs to be addressed.

#### 4.3.2 Diversity Issues

Our analyses suggest students perceive a significant lack of females in engineering. For example, "eighty five kids leaving the classroom before mine. of those 85, four are girls. Engineers math class #Stereotypical", and "Keeping up with tradition: 2 girls in a class of 40". This issue may be again related to the nerdy and anti-social image of engineering. Male students in engineering are regarded as bad at talking with female students, because they usually do not have many female students around in their class. For example, "I'm sorry. We're not use to having girls around", "I pity the 1 girl in my lab with 25 guys. It smells like man in here.... And that's not in a good way", "Finally talked to a girl today!!! It was Siri", and "Let's start with an example, tell me something you know nothing about' – Professor ... 'girls.' – Students. lol".

Another diversity issue is reflected by the complaints that there are too many foreign (or international) professors and students around. Students say that they cannot understand the accents of foreign professors in class, and they do not like to work with foreign students in course projects. Some examples are "I think its a requirement for engineering professors to be Asian and have terrible accents!", "not one of my professors can speak....?? #foreigners #mad accents", "When I signed on for engineering, I didn't realize I'd be getting a minor in foreign accents..."," Everywhere I go in this building it smells like Indian food and f ", and "Avoiding doing group work with international students." The issue here is not lack of diversity, but rather that students have difficulties embracing the diversity, because of many culture conflicts.

#### 4.3.3 Others: The Long Tail

A large number of tweets fall under this category. Many tweets in this category do not have a clear meaning. Other tweets in this category do reflect various issues that engineering students have but seen in very small volumes. Examples of these issues include curriculum problems, lack of motivation, procrastination, career and future worries, identity crisis, thought of switching majors, and physical health problems. Anderson elaborated the concept of "long tail" wherein e-commerce sites such as Amazon can maintain a large number of less popular products. Collectively these less popular items create more revenues than the best-sellers [50]. The concept of "long tail" is also used in social tagging systems and many other types of user-generated content on the Internet. In social tagging systems, a small number of prevalent tags appear with high frequency and a large number of unique tags fall into the "long tail" of a power law distribution [51], [52]. Ochoa and Duval [53] show that many other types of user-generated content online including reviews, photos, videos, and social news all follow "long tail" distributions.

In the case of our study, the tweets are user-generated content on social media. A small number of common student problems appear in high frequency, and a large number of less common problems or noisy tweets each appear in very low frequency. This indicates a "long tail" character. It is a challenge and also our future work to reveal more insightful information from this long tail.

## 5 NA€IVE BAYES MULTI-LABEL CLASSIFIER

We built a multi-label classifier to classify tweets based on the categories developed in the previous content analysis stage. There are several popular classifiers widely used in data mining and machine learning domain. We found Na€ıve Bayes classifier to be very effective on our data set compared with other state-of-the-art multi-label classifiers, which we compared in Section 6.

### 5.1 Text Pre-Processing

Twitter users use some special symbols to convey certain meaning. For example, # is used to indicate a hashtag, @ is used to indicate a user account, and RT is used to indicate a re-tweet. Twitter users sometimes repeat letters in words so that to emphasize the words, for example, "huuungryyy", "sooomuuchh", and "Monnndayyy". Besides, common Stop words such as "a, an, and, of, he, she, it", non-letter symbols, and punctuation also bring noise to the text. So we pre-processed the texts before training the classifier:

1. We removed all the #engineering Problems hashtags. For other co-occurring hashtags, we only removed the # sign, and kept the hashtag texts.
2. Negative words are useful for detecting negative emotion and issues. So we substituted words ending with "n't" and other common negative words (e.g., nothing, never, none, cannot) with "negtoken".

3. We removed all words that contain non-letter symbols and punctuation. This included the removal of @ and http links. We also removed all the RTs.
4. For repeating letters in words, our strategy was that when we detected two identical letters repeating, we kept both of them. If we detected more than two identical letters repeating, we replaced them with one letter. Therefore, "huuungryyy" and "sooo" were corrected to "hungry" and "so". "muuchh" was kept as "muuchh". Originally correct words such as "too" and "sleep" were kept as they were.
5. We used the Lemur information retrieval toolkit [54] to remove the common stopwords. We kept words like "much, more, all, always, still, only", because the tweets frequently use these words to express extent. The Krovetz stemmer in the Lemur toolkit was used to perform stemming in order to unify different forms of a word, such as plurals and different forms of a verb.

**5.2 Na€ıve Bayes Multi-Label Classifier**

One popular way to implement multi-label classifier is to transform the multi-label classification problem into multiple single-label classification problems [55]. One simple transformation method is called one-versus-all or binary relevance [55]. The basic concept is to assume independence among categories, and train a binary classifier for each category. All kinds of binary classifier can be transformed to multi-label classifier using the one-versus-all heuristic. The following are the basic procedures of the multi-label Na€ıve Bayes classifier. Suppose there are a total number of N words in the training document collection (in our case, each tweet is a document) W ¼ fw$_1$;w$_2$;:::;w$_N$g, and a total number of L categories C ¼ fc$_1$;c$_2$;:::;c$_L$g. If a word $w_n$ appears in a category c for $m_{wn}$ctimes, and appear in categories other than c for $m_{wn}$c times, then based on the maximum likelihood estimation, the probability of this word in a specific category c is

$$P\left(w_n\left|c\right.\right)=\frac{N}{P_n^{\frac{1}{4}}1\ mw_n c} \quad (2)$$

Similarly, the probability of this word in categories other than c is

$$P\left(w_n\left|c\right.\right)=m\left(\frac{N}{P_n^{\frac{1}{4}}1\ mw_n c}\right) \quad (3)$$

Suppose there are a total number of M documents in the training set, and C of them are in category c. Then the probability of category c is

$$P\left(c\right)=\frac{C}{M} \quad (4)$$

and the probability of other categories c' is

$$P\left(w_n\left|c\right.\right)=P_n^{\frac{1}{4}}1\ mw_n c \quad (5)$$

For a document $d_i$ in the testing set, there are K wordsW$_{di}$¼ fw$_{i1}$; wi2;...;w$_{i K}$g, and W$_{di}$ is a subset of W. The purpose is to classify this document into category c or not c. We assume independence among each word in this document, and any word $w_{ik}$ conditioned on c or $c^0$ follows multinomial distribution. Therefore, according to Bayes' Theorem, the probability that $d_i$ belongs to category c is And the probability that $d_i$ belongs to categories other than c Is

$$P\left(w_n\left|c\right.\right)=P_n^{\frac{1}{4}}1 \quad (6)$$

Because pðcjd$_i$Þ þ pðc$^0$ jd$_i$Þ ¼ 1, we normalize the latter two items which are proportional to pðcjd$_i$Þand pðc$^0$ jd$_i$Þto get the real values of pðcjd$_i$Þ. If pðcjd$_i$Þis larger than the probability threshold T, then $d_i$ belongs to category c, otherwise, $d_i$ does belong to category c. Then repeat this procedure for each category. In our implementation, if for a certain document, there is no category with a positive probability larger than T, we assign the one category with the largest probability to this document. In addition, "others" is an exclusive category. A tweet is only assigned to "others" when "others" is the only category with probability larger than T. Section 5.4 details the choices of the threshold values T.

**5.3 Evaluation Measures for Multi-Label Classifier**

Commonly used measures to evaluate the performance of classification models include accuracy, precision, recall, and the harmonic average between precision and recall—the $F_1$ score. For multi-label classification, the situation is slightly more complicated, because each document gets assigned multiple labels. Among these labels, some may be correct, and others may be incorrect. Therefore, there are usually two types of evaluation measures—example-based measures and label-based measures

[55]. Example-based measures are calculated on each document (e.g., each tweet is a document, and also called an example here) and then averaged over all documents in the data set, whereas label-based measures are calculated based on each label (category) and then averaged over all labels (categories).

### *5.3.1 Example-Based Evaluation Measures*

For a certain document d, suppose the true set of labels it falls under is Y , and the predicted set of labels Contingency Table per Category

Table 1

| | True *c* | True not *c* |
|---|---|---|
| Predicted *c* | true positive (*tp*) | false positive (*fp*) |
| Predicted not *c* | false negative (*fn*) | true negative (*tn*) |

*The sum of tp, fp, fn, and tn equal to the total number of documents.*

TABLE 2

| *Category* | *Label a.* | *Label $F_1$* | *Rand. a.* | *Rand. $F_1$* |
|---|---|---|---|---|
| *Heavy Study Load* | 0.8514 | 0.4851 | 0.0756 | 0.0136 |
| *Lack of Social Engagement* | 0.9165 | 0.5072 | 0.0821 | 0.0099 |
| *Negative Emotion* | 0.9644 | 0.5039 | 0.0877 | 0.0062 |
| *Sleep Problems* | 0.9595 | 0.6916 | 0.0851 | 0.0077 |
| *Diversity Issues* | 0.9226 | 0.6882 | 0.0837 | 0.0088 |
| *Others* | 0.7432 | 0.8107 | 0.0527 | 0.0617 |

*Label a. = label-based accuracy (12), Label $F_1$ = label-based $F_1$ (15),*
*Rand. a. = random guessing label-based accuracy,*
*Rand. $F_1$ = random guessing label-based $F_1$.*

Six Evaluation Measures with Na€ıve Bayes Multi-Label Classifier under Different Probability Thresholds classifier is Z, then for this specific document, accuracy is the correctly predicted number of labels divided by the number of labels in the union of Y and Z. Precision is the correctly predicted number of labels divided by the total number of labels in Z, while recall is the correctly predicted number of labels divided by the number of true labels. Suppose there are a total of M documents fd$_1$;d$_2$;...;d$_M$g, the accuracy, precision, recall, and $F_1$ averaged over the M documents.

### *5.3.2 Label-Based Evaluation Measures*

Rather than calculated and averaged over each document instance like in the example-based measures, label-based measures are calculated and averaged over each category. In each of the one-versus-all binary classification step, we can create the matrix in Table 1 for the corresponding category c. Micro-averaging gives equal weight to each per-document classification decision, while macro-averaging gives equal weight to each category [56]. Thus micro-averaging score is dominated by categories that have larger number of documents, while macro-averaged $F_1$ is closer to the algorithm effectiveness on smaller categories. So micro-averaged $F_1$ is higher for classifiers work well on large categories, while macro-averaged $F_1$ is higher for classifiers work better on smaller categories. Actually, equations (1), (11), (15), (16), and (17) are variations of $F_1$ used in different situations.

### 5.4 Classification Results

From the inductive content analysis stage, we had a total of 2,785 #engineering Problems tweets annotated with 6 categories. We used 70 percent of the 2,785 tweets for training (1,950 tweets), and 30 percent for testing (835 tweets). 85.5 percent (532/622) of words occurred more than once in the testing set were found in the training data set. Table 2 shows the six evaluation measures at each probability threshold values from 0 to 1 with a segment of 0.1. We assigned the one category with the largest probability value to the document when there was no category with a positive probability value larger than T. So when the probability threshold was 1, it was equivalent to outputting the largest possible one category for all the tweets. With five multi-label categories and one “others” category, there are ð2$^5$ 1Þ þ 1 ¼ 32 possible label sets for a 3 Label-Based Accuracy and $F_1$ for Each Category Naıve Bayes versus Random Guessing

Tweet. Tables 2 and 3 provide all the evaluation measures under random guessing. The random guessing program first guessed whether a tweet belongs to “others” based on the proportion this category takes in the training data set. If this tweet did not belong to “others”, it then proceeded to guess whether it fell into the rest of the categories also based the proportion each category takes in the rest categories. We repeated the random guessing program 100 times, and obtained the average measures. From Table 2, we see that when the probability threshold value is 0:4, the performance is generally better than under other threshold values. Table 3 shows the label-based accuracy (12) and $F_1$ measure (15) for each of the six category when T ¼ 0:4 compared with random guessing. Based on (12), label-based accuracy accounts for true negative (tn) numbers. Therefore, under the situation that there are a large number of tweets in “others”, accuracy measures are much higher than $F_1$ scores in Table 3. Label based accuracy is not a very effective measure to account label imbalance here, so we do not use this measure in further discussion. The Na€ıve Bayes classifier has not only achieved significant improvement from the random guessing baseline, but also exceeded the performance of state-of-the-art multi-label classifiers on our data set as shown below.

**6 COMPARISON EXPERIMENT: SVM AND M3L**

Support Vector Machine (SVM) [57] is one of the most used and accurate classifiers in many machine learning tasks, but our comparison experiment shows that Naıve Bayes exceeds SVM in this study. We first implemented a linear multi-label SVM using the LibSVM library [58] with the one-versus-all heuristic. We applied weight of loss parameters that are proportional to the inverse of the

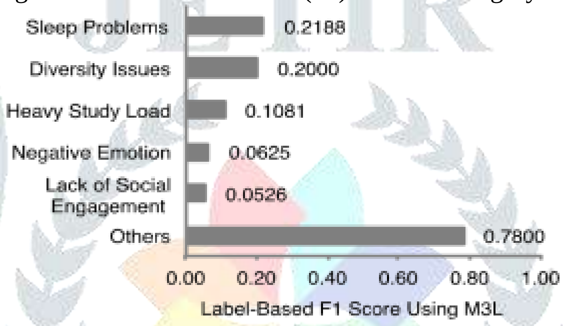
TABLE 4

Six Evaluation Measures with M3L

| *Ex. a* | *Ex. p* | *Ex. r* | *Ex. $F_1$* | *Micro. $F_1$* | *Macro. $F_1$* |
|---|---|---|---|---|---|
| 0.6384 | 0.6384 | 0.6388 | 0.6386 | 0.6339 | 0.2370 |

*Ex. a = example-based accuracy (8), Ex. p = example-based precision (9), Ex. r = example-based recall (10), Ex. $F_1$ = example-based $F_1$ (11), Micro. $F_1$ = micro-averaged $F_1$ (16), Macro $F_1$ = macro-averaged $F_1$ (17).*

Fig. 3. Label-based F1 score (15) for each category us




percentages of tweets in or not in each category to account for the imbalanced categories. However, with the same training and testing data sets as in the above section, this one-versus-all SVM multi-label classifier classified all tweets into not in the category for all categories. So we got empty label sets for all tweets.

Then we applied the same training and testing data sets as above to an advanced SVM variation named Max Margin Multi-Label classifier. M3L is a state-of-the-art multi-label classifier [59]. Different from the one-versus all heuristic, which assumes label independence, this classifier takes label correlation into consideration. We used the executable file of this algorithm provided by the authors [60]. The performance is better than the simplistic one-versus-all SVM classifier, but still not as good as the Naive Bayes classifier. Table 4 and Fig. 3 show the evaluation measures using M3L. Because SVM is not a probabilistic model, so Table 4 does not have probability threshold values as Table 2 does.

**7 DETECT STUDENT PROBLEMS FROM PURDUE DATA SET**

In this section, the Naıve Bayes multi-label classifier is used to detect engineering student problems from the Purdue data set. There were 35,598 unique tweets in the Purdue tweet collection. We took a random sample of 1,000 tweets, and found no more than 5 percent of these tweets were discussing engineering problems. Our purpose here was to detect the small number of tweets that reflect engineering students’ problems. The differences between #engineering Problems data set and Purdue data set is that the latter contains much smaller number of positive samples to be detected, and its “others” category has more diverse content. Therefore, to make the training set better adapt to the Purdue data set, we took a random sample of 5,000 tweets from the Purdue data set, added them into the 2,785 #engineering Problems tweets, and labeled them as “others”. Less than 5 percent positive samples in this category do not influence the effectiveness of the trained model. We thus used the 7,785 tweets as input to train the multi-label Naıve Bayes classifier. Since no extra human effort is needed, and Naive Bayes classifier is very efficient in terms of computation time, the model training here incurred almost no extra cost. Table 5 shows the top most probable words in each category ranked using the conditional probability p(wjc)Þ as in (2). Our purpose here is to detect the small number of the five problems from the large Purdue data set, so we do not discuss the “others” in this section.

5 Top 25 Most Probable Words in Each Category

| *Category* | *Top 25 words* |
|---|---|
| ***Heavy Study Load*** | hour, homework, exam, day, class, work, negtoken, problem, study, week, toomuch, all, lab, still, out, time, page, library, spend, today, long, school, due, engineer, already |

| | |
|---|---|
| *Lack of Social Engagement* | negtoken, Friday, homework, out, study, work, weekend, life, class, engineer, exam, drink, break, Saturday, people, social, lab, spend, tonight, watch, game, miss, party, sunny, beautiful |
| *Negative Emotion* | hate, f***, shit, exam, negtoken, week, class, hell, engineer, suck, study, hour, homework, time, equate, FML, lab, sad, bad, day, feel, tired, damn, death, hard |
| *Sleep Problems* | sleep, hour, night, negtoken, bed, allnight, exam, homework, nap, coffee, time, study, more, work, class, dream, ladyengineer, late, week, day, long, morning, wake, awake, no-sleep |
| *Diversity Issues* | girl, class, only, negtoken, guy, engineer, Asia, professor, speak, English, female, hot, kid, more, toomuch, walk, people, teach, understand, chick, China, foreign, out, white, black |

We can see that each category has top words for the specific content of this category. For example, "weekend, life, drink, social, break" for the category Lack of Social Engagement, "hate, suck, sad, bad" for the category Negative Emotion, "sleep, all night, nap, coffee, dream" for Sleep Problems, and "girl, guy, Asia, female, China, foreign" for Diversity Issues. This intuitively demonstrates the effectiveness of the classification model.

We applied the trained model to the rest 30,598 Purdue tweets, and detected a total of 940 tweets reflecting the five problems students encounter shown in Fig. 4. Again, one tweet may fall under several different categories, so the sum of numbers of tweets of all categories appears more than 940. We see a relatively small number of tweets reflecting the diversity issues, which is not as severe as reflected by the #engineering Problems tweets. The College of Engineering at Purdue has a large number of international students, and is constantly making efforts to increase the enrollment of

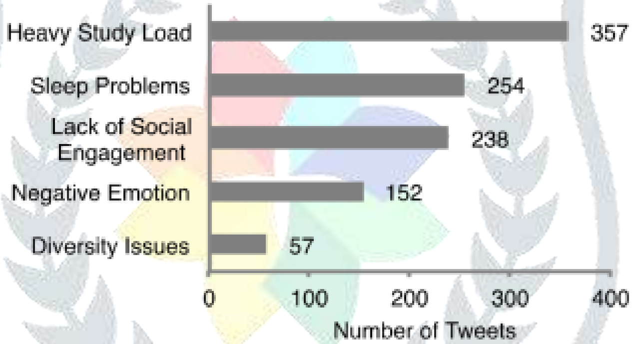



Fig. 4. Number of tweets for each issue detected from the Purdue tweet collection. We do not discuss "others", because we detect tweets reflecting these five problems from large number of Purdue tweets.

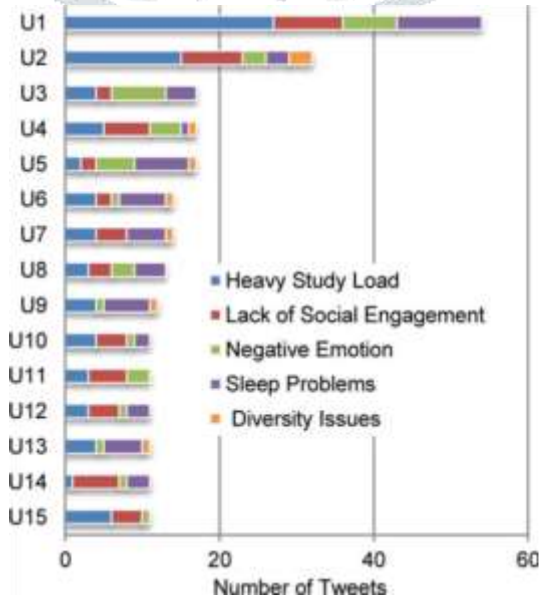


Fig. 5. Top 15 users in the Purdue tweet collection who posted the most on the five engineering problems. Table 6 shows the total number of tweets these users posted under the five categories, and the percentages tweets in these five categories take in the entire number of tweets posted by these users.

female and students in other underrepresented groups. It is a hypothesis that there is a correlation between Purdue's effort and what the data show. Each of the 940 tweets is associated with a Twitter user account. In Fig. 5, we show the top 15 users that post the most of the five engineering problems. The twitter user names were anonymized to protect their privacy. 940 tweets take 3.07 percent of the 30,598 tweets. As shown in Table 6, many of the top 15 users have posted more than 3.07 percent tweets on the five engineering problems.

TABLE 6
Top 15 Users' Numbers of Tweets on the Five Engineering Problems and Percentages Taken in the Total Number of Tweets Posted

| User | Tweets on the Five Eng. Problems | All Tweets Posted | Percentage |
|---|---|---|---|
| U1 | 54 | 669 | 8.07% |
| U2 | 32 | 279 | 11.47% |
| U3 | 17 | 206 | 8.25% |
| U4 | 17 | 330 | 5.15% |
| U5 | 17 | 642 | 2.65% |
| U6 | 14 | 171 | 8.19% |
| U7 | 14 | 203 | 6.90% |
| U8 | 13 | 246 | 5.28% |
| U9 | 12 | 341 | 3.52% |
| U10 | 11 | 110 | 10.00% |
| U11 | 11 | 114 | 9.65% |
| U12 | 11 | 133 | 8.27% |
| U13 | 11 | 179 | 6.15% |
| U14 | 11 | 285 | 3.86% |
| U15 | 11 | 465 | 2.37% |

7
Evaluation Measures on Purdue Data Set

| *Ex. a* | *Ex. p* | *Ex. r* | *Ex.* $F_1$ |
|---|---|---|---|
| 0.5519 | 0.6057 | 0.5829 | 0.5799 |

*Ex. a = example-based accuracy (8), Ex. p = example-based precision (9), Ex. r = example-based recall (10), Ex.* $F_1$ *= example-based* $F_1$ *(11). Micro-averaged* $F_1$ *and macro-averaged* $F_1$ *are not provided here, because their calculation requires the ground truth of the entire 30,598 tweets, which is impossible to know.*

Fig. 5 and Table 6 demonstrate that our approach has the ability to detect potential student problems from tweets. Yet, we are not trying to make the claim that these users shown are definitely at-risk students, since most of these users only posted less than 10 percent of problems among all tweets they have posted. The trained detector can be applied as a monitoring mechanism in the long run to identify severe cases of at-risk students. For example, a future student may post a large number of tweets and more than 90 percent of them are about study problems or negative emotions. This case can be detected using the detector presented here. Our Purdue data set only lasts a little over two months. However, severe cases may only appear once in a while depending on the institutional atmosphere. Further decisions need to be made about what counts as severe cases, to what extent does intervention is needed, how to protect students' privacy, and how comfortable they are about these interventions.

To further evaluate the performance of the Na€ıve Bayes classifier on the Purdue data set, we manually checked the 940 tweets and the corresponding user accounts. Since it is impossible to manually check all 30,598 tweets, we only provide the example-based measurements and the precisions for each category as in Table 7 and Fig. 6. We did not run the random guessing program here, because as shown in Table 2, example-based measures resulting from random guessing on 32 possible label sets is smaller than 0.05. In addition, it can be easily proved that for label-based precision measure (13) as in Fig. 6, the random guessing precision for any category equals to the actual number of tweets in this category divided by the total number of tweets in the entire collection. In this case, there are less than 5 percent of tweets in the Purdue data set that fall into the five engineering student problems. Therefore, random guessing precision is smaller than 0.05. As of now, the performance of the detector is not superior, but has achieved significant improvement from random guessing baseline. As illustrated in the workflow in Fig. 1, the performance of the algorithm can be gradually improved based on further human feedback.

## 8 DISCUSSION, LIMITATIONS, AND FUTURE WORK

This study explores the previously un instrumented space on Twitter in order to understand engineering students' experiences, integrating both qualitative methods and large-scale data mining techniques. In our study, through a qualitative content analysis, we found that engineering students are largely struggling with the heavy study load, and are not able to manage it successfully.

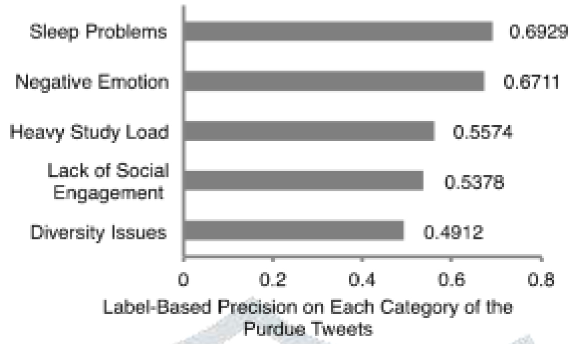



Fig. 6. Label-Based precisions (13) on each category of the 940 Purdue tweets detected from 30,598 tweets using Na€ıve Bayes. Unlike in Table 4 and Fig. 3, $F_1$ score is impossible to calculate because the ground truth of the entire 30,598 tweets is unknown.

Heavy study load leads to many consequences including lack of social engagement, sleep problems, and other psychological and physical health problems. Many students feel engineering is boring and hard, which leads to lack of motivation to study and negative emotions. Diversity issues also reveal culture conflicts and culture stereotypes existing among engineering students.
Building on top of the qualitative insights, we implemented and evaluated a multi-label classifier to detect engineering student problems from Purdue University. This detector can be applied as a monitoring mechanism to identify at-risk students at a specific university in the long run without repeating the manual work frequently.

Our work is only the first step towards revealing actionable insights from student-generated content on social media in order to improve education quality. There are a number of limitations, which also lead to many possible directions for future work. First, not all students are active on Twitter, so we may only find the ones who are more active and more likely to expose their thoughts and feelings. Also, students' awareness of identity management online may increase overtime. The "manipulation" of personal image online may need to be taken into considerations in future work. Second, the fact that the most relevant data we found on engineering students' learning experiences involve complaints, issues, and problems does not mean there is no positive aspects in students' learning experiences. We did find a small number of tweets that discuss the good things about being engineering students such as those using hashtag #engineering Perks. We chose to focus on the problems in this paper because these could be the most informative for improvement of education quality. Future work can compare both the good and bad things to investigate the tradeoffs with which students struggle. From another aspect, students tend to complain about issues and problems on social media. This may imply that social media serve as a good venue for students to vent negative emotions and seek social support. Therefore, future work can be done on why and how students seek social support on social media sites.

Third, we only identified the prominent themes with relatively large number of tweets in the data. There are a variety of other issues hidden in the "long tail". Several of these issues may be of great interest to education researchers and practitioners. Future work can be done to design more sophisticated algorithms in order to reveal the hidden information in the "long tail". Fourth, the qualitative analysis reveals that there are correlations among the themes. For example, "heavy study load" can cause "lack of social engagement" and "sleep problems". Also, "negative emotion" may be associated with several other themes. The Na€ıve Bayes classifier is built on top of the label independence assumption, which is a simplification of the real world problem. The classifier is designed to be a multi-label classifier in order to reconcile this effect. If a tweet expresses correlation between "heavy study load" and "sleep problems", then it can be categorized into both categories. After all, any mathematical and statistical models are simplification of real world problems to a certain extent. The comparison experiment with M3L shows that this advanced model that accounts for label correlation does not perform as well as the simple Na€ıve Bayes model. Future work could specifically address the correlations among these student problems.

Finally, the workflow we proposed requires human effort for data analysis and interpretation. This is necessary because our purpose is to achieve deeper understanding of the student experiences. To the best of our knowledge, there is currently no unsupervised automatic natural language processing technique that can achieve the depth of understanding that we were able to achieve. There is a tradeoff between the amount of human effort and the depth of the understanding. The labels generated can be applied to any similar data sets in other institutions to detect engineering student problems without extra human effort. Often times, manual analysis is time-consuming not only because of the time spent on analyzing the actual data, but also the time spent on cleaning, organizing the data, and adapting the format to fit the algorithms. We plan to build a tool based on the workflow proposed here combining social media data and possibly student academic performance data. This tool can assist in identification of students at risk. This tool will provide a friendly user interface and integration between qualitative analysis and the classification and detection algorithms [61]. Therefore, educators and researchers using this tool can focus on the actual data analysis and investigate the types of learning issues that they perceive as critical to their institutions and students. This tool can also facilitate collaboration among researchers and educators on data analysis. Advanced natural language processing techniques can be applied in the future to provide topic recommendations and further augment the human analysis results, but cannot completely rule out the human effort.

Other possible future work could analyze students' generated content other than texts (e.g., images and videos), on social media sites other than Twitter (e.g., Facebook, Tumbler, and YouTube). Future work can also extend to students in other majors and other institutions.

**9 CONCLUSION**

Our study is beneficial to researchers in learning analytics, educational data mining, and learning technologies. It provides a workflow for analyzing social media data for educational purposes that overcomes the major limitations of both manual qualitative analysis and large scale computational analysis of user-generated textual content. Our study can inform educational administrators, practitioners and other relevant decision makers to gain further understanding of engineering students' college experiences. As an initial attempt to instrument the uncontrolled social media space, we propose many possible directions for future work for researchers who are interested in this area. We hope to see a proliferation of work in this area in the near future. We advocate that great attention needs to be paid to protect students' privacy when trying to provide good education and services to them.

**10 REFERENCES**


[1] G. Siemens and P. Long, "Penetrating the Fog: Analytics in Learning and Education," Educause Rev., vol. 46, no. 5, pp. 30-32, 2011.

[2] M. Rost, L. Barkhuus, H. Cramer, and B. Brown, "Representation and Communication: Challenges in Interpreting Large Social Media Datasets," Proc. Conf. Computer Supported Cooperative Work, pp. 357-362, 2013.

[3] M. Clark, S. Sheppard, C. Atman, L. Fleming, R. Miller, R. Stevens, R. Streveler, and K. Smith, "Academic Pathways Study: Processes and Realities," Proc. Am. Soc. Eng. Education Ann. Conf. Exposition, 2008.

[4] C.J. Atman, S.D. Sheppard, J. Turns, R.S. Adams, L. Fleming, R. Stevens, R.A. Streveler, K. Smith, R. Miller, L. Leifer, K. Yasuhara, and D. Lund, Enabling Engineering Student Success: The Final Report for the Center for the Advancement of Engineering Education. Morgan & Claypool Publishers, Center for the Advancement of Engineering Education, 2010.

[5] R. Ferguson, "The State of Learning Analytics in 2012: A Review and Future Challenges," Technical Report KMI-2012-01, Knowledge Media Inst. 2012.

[6] R. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," J. Educational Data Mining, vol. 1, no. 1, pp. 3-17, 2009.

[7] S. Cetintas, L. Si, H. Aagard, K. Bowen, and M. Cordova-Sanchez, "Microblogging in Classroom: Classifying Students' Relevant and Irrelevant Questions in a Microblogging-Supported Classroom," IEEE Trans. Learning Technologies, vol. 4, no. 4, pp. 292-300, Oct.Dec. 2011.

[8] C. Moller-Wong and A. Eide, "An Engineering Student Retention Study," J. Eng. Education, vol. 86, no. 1, pp. 7-15, 1997.

[9] Nat'l Academy of Eng., The Engineer of 2020: Visions of Engineering in the New Century. National Academies Press, 2004.

[10] E. Goffman, The Presentation of Self in Everyday Life. Lightning Source Inc., 1959.

[11] E. Pearson, "All the World Wide Web's a Stage: The Performance of Identity in Online Social Networks," First Monday, vol. 14, no. 3, pp. 1-7, 2009.

[12] J.M. DiMicco and D.R. Millen, "Identity Management: Multiple Presentations of Self in Facebook," Proc. the Int'l ACM Conf. Supporting Group Work, pp. 383-386, 2007.

[13] M. Vorvoreanu and Q. Clark, "Managing Identity Across Social Networks," Proc. Poster Session at the ACM Conf. Computer Supported Cooperative Work, 2010.

[14] M. Vorvoreanu, Q.M. Clark, and G.A. Boisvenue, "Online Identity Management Literacy for Engineering and Technology Students," J. Online Eng. Education, vol. 3, article 1, 2012.

[15] M. Ito, H. Horst, M. Bittanti, D. boyd, B. Herr-Stephenson, P.G. Lange, S. Baumer, R. Cody, D. Mahendran, K. Martinez, D. Perkel, C. Sims, and L. Tripp, Living and Learning with New Media: Summary of Findings from the Digital Youth Project. The John D. and Catherine T. MacAuthur Foundation, Nov. 2008.

[16] D. Gaffney, "#iranElection: Quantifying Online Activism," Proc. Extending the Frontier of Society On-Line (WebSci10), 2010.

[17] S. Jamison-Powell, C. Linehan, L. Daley, A. Garbett, and S. Lawson, "'I Can't Get No Sleep': Discussing #Insomnia on Twitter," Proc. ACM Ann. Conf. Human Factors in Computing Systems, pp. 1501-1510, 2012.

[18] M.J. Culnan, P.J. McHugh, and J.I. Zubillaga, "How Large US Companies Can Use Twitter and Other Social Media to Gain Business Value," MIS Quarterly Executive, vol. 9, no. 4, pp. 243-259, 2010.

[19] M.E. Hambrick, J.M. Simmons, G.P. Greenhalgh, and T.C. Greenwell, "Understanding Professional Athletes' Use of Twitter: A Content Analysis of Athlete Tweets," Int'l J. Sport Comm., vol. 3, no. 4, pp. 454-471, 2010.

[20] D.M. Romero, B. Meeder, and J. Kleinberg, "Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter," Proc. 20th Int'l Conf. World Wide Web, pp. 695-704, 2011.

[21] J. Yang and S. Counts, "Predicting the Speed, Scale, and Range of Information Diffusion in Twitter," Proc. Fourth Int'l AAAI Conf. Weblogs and Social Media (ICWSM), 2010.

[22] M. Gomez Rodriguez, J. Leskovec, and B. Scholkopf,€"Structure and Dynamics of Information Pathways in Online Media," Proc. Sixth ACM Int'l Conf. Web Search and Data Mining, pp. 23-32, 2013.

[23] R. Bandari, S. Asur, and B.A. Huberman, "The Pulse of News in Social Media: Forecasting Popularity," Proc. Int'l AAAI Conf. Weblogs and Social Media (ICWSM), 2012.

[24] L. Hong, O. Dan, and B.D. Davison, “Predicting Popular Messages in Twitter,” Proc. 20th Int’l Conf. Companion on World Wide Web, pp. 57-58, 2011.

[25] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors,” Proc. 19th Int’l Conf. World Wide Web, pp. 851-860, 2010.

[26] H. Becker, M. Naaman, and L. Gravano, “Selecting Quality Twitter Content for Events,” Proc. Fifth Int’l AAAI Conf. Weblogs and Social Media (ICWSM ’11), 2011.

[27] W. Zhao, J. Jiang, J. Weng, J. He, E.P. Lim, H. Yan, and X. Li, “Comparing Twitter and Traditional Media Using Topic Models,” Proc. 33rd European Conf. Advances in Information Retrieval, pp. 338349, 2011.

[28] L. Hong and B.D. Davison, “Empirical Study of Topic Modeling in Twitter,” Proc. First Workshop Social Media Analytics, pp. 80-88, 2010.

[29] D. Davidov, O. Tsur, and A. Rappoport, “Enhanced Sentiment Learning Using Twitter Hashtags and Smileys,” Proc. 23rd Int’l Conf. Computational Linguistics: Posters, pp. 241-249, 2010.

[30] A. Go, R. Bhayani, and L. Huang, “Twitter Sentiment Classification Using Distant Supervision,” CS224N Project Report, Stanford pp. 1-12, 2009.

[31] K. Nishida, R. Banno, K. Fujimura, and T. Hoshide, “Tweet Classification by Data Compression,” Proc. Int’l Workshop DETecting and Exploiting Cultural DiversiTy on the Social Web, pp. 29-34, 2011.

[32] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, “Short Text Classification in Twitter to Improve Information Filtering,” Proc. 33rd Int’l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 841-842, 2010.

[33] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs Up?: Sentiment Classification Using Machine Learning Techniques,” Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing, vol. 10, pp. 79-86, 2002.

[34] G. Siemens and R.S. d Baker, “Learning Analytics and Educational Data Mining: Towards Communication and Collaboration,” Proc. Second Int’l Conf. Learning Analytics and Knowledge, pp. 252254, 2012.

[35] K.E. Arnold and M.D. Pistilli, “Course Signals at Purdue: Using Learning Analytics to Increase Student Success,” Proc. Second Int’l Conf. Learning Analytics and Knowledge, pp. 267-270, 2012.

[36] R. Baker and K. Yacef, “The State of Educational Data Mining in 2009: A Review and Future Visions,” J. Educational Data Mining, vol. 1, no. 1, pp. 3-17, 2009.

[37] C. Romero and S. Ventura, “Educational Data Mining: A Review of the State of the Art,” IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Rev., vol. 40, no. 6, pp. 601-618, Nov. 2010.

[38] “Using the Twitter Search API | Twitter Developers,” https:// dev.twitter.com/docs/using-search, 2013.

[39] “Facts and Figures—College of Engineering, Purdue University,” https://engineering.purdue.edu/Engr/AboutUs/FactsFigures, 2013.

[40] D.M. Blei and J. D. Lafferty, “A Correlated Topic Model of Science,” The Annals of Applied Statistics, vol. 1, no. 1, pp. 17-35, 2007.

[41] Y.-C. Wang, M. Burke, and R.E. Kraut, “Gender, Topic, and Audience Response: An Analysis of User-Generated Content on Facebook,” Proc. SIGCHI Conf. Human Factors in Computing Systems, pp. 31-34, 2013.

[42] K. Krippendorff, “Reliability in Content Analysis,” Human Comm. Research, vol. 30, no. 3, pp. 411-433, 2004.

[43] M. Lombard, J. Snyder-Duch, and C.C. Bracken, “Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability,” Human Comm. Research, vol. 28, no. 4, pp. 587604, 2006.

[44] J.L. Devore, Probability & Statistics for Engineering and the Sciences. Duxbury Press, 2012.

[45] H. Loshbaugh, T. Hoeglund, R. Streveler, and K. Breaux, “Engineering School, Life Balance, and the Student Experience,” Proc. ASEE Ann. Conf. & Exposition, 2006.

[46] S. Cohen and T. A. Wills, “Stress, Social Support, and the Buffering Hypothesis,” Psychological Bull., vol. 98, no. 2, pp. 310-357, 1985.

[47] A. Pritchard and J. Woollard, Psychology for the Classroom: Constructivism and Social Learning. Taylor & Francis, 2010.

[48] H. Loshbaugh and B. Claar, “Geeks Are Chic: Cultural Identity and Engineering Students’ Pathways to the Profession,” Proc. Am. Soc. for Eng. Education (ASEE) Conf., 2007.

[49] ABET, “ABET—Criteria for accrediting engineering programs, 2012—2013,” http://www.abet.org/DisplayTemplates/ Docs Handbook.aspx?id=3143, 2012-2013.

[50] C. Anderson, The Long Tail: Why the Future of Business Is Selling Less of More. Hyperion Books, 2008.

[51] E. Tonkin, “Searching the Long Tail: Hidden Structure in Social Tagging,” Proc. 17th ASIS&SIG /CR Classification Research Workshop, 2006.

[52] S.A. Golder and B.A. Huberman, “The Structure of Collaborative Tagging Systems,” J. Information Science, vol. 32, no. 2, pp. 198-208, 2006.

[53] X. Ochoa and E. Duval, “Quantitative Analysis of User-Generated Content on the Web,” Proc. First Int’l Workshop Understanding Web Evolution (webevolve ’08), 2008.

[54] “Lemur Information Retrieval Toolkit,” http://www. lemurproject.org/,2013.