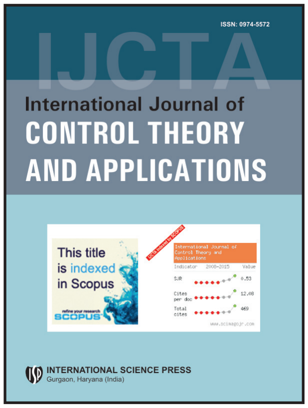


**International Journal of Control Theory and Applications**

ISSN : 0974-5572



**Volume 9 • Number 43 • 2016**

# Deep Learning Analysis of Engineering Students' Emotions through Social Media Data Mining

**Akshay D. Patil[a] and Umesh B. Chavan[b]**

[a]*P.G. Student, Department of Computer Science and Engineering, Walchand College of Engineering, Vishrambag, Sangli, Maharashtra, India. Email: akshaypatil6362@gmail.com*
[b]*Assistant Professor, Department of Information Technology, Walchand College of Engineering, Vishrambag, Sangli, Maharashtra, India. Email: umesh.chavan@walchandsangli.ac.in*

***Abstract:*** Engineering students casual daily discussion on social media focused into their educational life experiences, mindset, and worries about the learning procedure. Observing and analyzing data from such a social media can be challenging work. The problem of students experiences reveal from social media sited requires human interpretation analysis or Interaction. It pays attention on engineering students Twitter posts to know such problems in their educational experiences with practices. This paper proposes a flow of work to put together both qualitative analysis and large-scale data mining techniques. First a sample tweets is taken from student and then qualitative analysis conducted with human interpretation on that sample which is associated to engineering students educational life experiences. So only tweets related to engineering student problems are collected. It is found that engineering students encounter problems such as Stressful, WorkLoad, Tension, Tough, Missing Social Engagement, Negativity, Sadness, Boring, Less Time Sleep, Major Diversion, Struggle. Based on this collection, a multi-label classification algorithm that is Naive Bayes Multi-label Classifier algorithm is applied to categorize tweets presenting students problems[1]. The algorithm prepares a train detector of engineering student problems.

***Keywords:*** Social Media Data Analysis, Educational Data Mining, Qualitative Data Analysis, Large-Scale Data Mining Techniques.

## 1. INTRODUCTION

Data mining research study has effectively produced several techniques, tools, schemes and algorithms for managing and analyzing large amounts of data to response real-world troubles. As social networking media is used for so many purposes, very vast amounts of user created data be present and can be made available for data mining. Mining the data of social network media can enlarge researchers ability of understanding innovative experiences in education life, to the use of social medium to represent better services and extend innovative opportunities. Main objectives of the data mining procedure are to collectively handle large amount of data, extract valuable, understandable and actionable patterns, and gain insightful knowledge. Social networking media sites such as Twitter, Facebook, and YouTube present best platform to students to share happiness and struggle, sentiment and tension, and gain social support helpfully.

On so many social networking media sites, students speak about their everyday encountered issue s in a comfortable and informal manner. This students valuable digital information gives large amount of implicit data information and a overall new viewpoint idea for educational researchers to know students valuable experiences outside the prohibited classroom environment. This useful understanding can enhance education system quality, and hence it improves students employment, preservation, and achievement. The huge amount of information on social media sites provides prospective to recognize and analyze students problem, but it increases some practical methodological complexities in use of social networking media sites data for educational purposes. The complexities such as absolute data volumes, the miscellany of internet slangs, the variation between locations, and different moments of students posting on the social media sites. Pure physical analysis can't deal with the ever growing large scale of data, while pure automatic algorithms can't capture in-depth significance inside the data or information.

A. *Research Gaps in Existing System:*

1. In Traditional system, educational researchers have been using methods such as interviews, surveys, focus groups, and classroom activities to collect information related to students education learning experiences.
2. These all above methods are usually very time-consuming, hence can't be duplicated or repeated with high frequency.
3. The scale of such studies is also usually limited and restricted.
4. Also when prompted about their learning experiences, students need to reflect on what they were thinking and doing something in the past, which may have become obscured over the time.
5. The emerging fields of learning analytics and educational data mining (EDM) have focused on analyzing such structured data produced from Course Management Systems (CMS), classroom based technology usage, or controlled online learning environments to inform educational researchers for decision-making.
6. However as per our knowledge, there is no research found to directly mine and analyze students posted contents from uncontrolled spaces on the social media with the clear goal of understanding students education learning experiences.

B. *Research Goals of this Study:*

1. To present a work of social media information sense-making for educational purposes, combining both qualitative analysis or investigation and large-scale data mining schemes.
2. To find engineering students informal casual discussions on Twitter, in manner to knows the problems coming into their engineering education life.

C. *Significance of this Research:* The Significance of this research study is prefers to focus on engineering students opinions posted on Twitter about their problems in college life because of:

1. Engineering institutes and respective branches have long been struggled with students employment and preservation chapters. Engineering graduates becomes a very significant part of the nations potential labor force and have a directly impact on the nations economical expansion.

2. Based on understanding of engineering students problems decision makers can make decisions on proper interference that can help students to combine obstacles in educational life and it will helpful the student to solve such problems.

3. Twitter is a well popular social networking media site. It's contents is recently updating public and very important thing that is no more than 140 characters per tweet. Twitter offers most useful free APIs that is used to stream data and allows developers to build upon and use their applications in new, innovative and creative ways. Access Data from Twitter give developers low latency access to Twitters global stream of Tweet data. To make a data mining design or are included in analytical research, the streaming API is most suitable use for such things. Twitter facts are in a proper suitable format for investigation. Twitters terms of apply for the data are relatively tolerant. It is generally accepted that tweets are public and available to anyone freely; hence they by default grant permission enter to any account with no need to request for sanction.

## 2. RELATED WORK

A. *Social Network Analysis:* The concept of social networks, typically seen as interactions between people, has become very much popular in the last decade due to its biggest application in the online domain. Although the structure of such social networks can be represented in terms of a graphs, plots or patterns etc., they often enclose huge amounts of information in terms of the linkage between people and contents shared among them[4].

B. *Educational Data Mining:* Educational data mining process converts data coming from different educational systems, such as traditional classroom environments, e-learning and intelligent learning systems into information that may be useful for educational researchers, teachers, professors, institutions and students on understanding and evaluating educational systems, aiming for improving the quality of the educational system process. For that purpose, the educational data mining process aims for developing and applying computational approaches such as data mining, web mining, and statistical schemes, for allowing the automatic information extraction from large amounts of data [7][8].

C. *Qualitative Research Analysis:* Qualitative research analysis is a technique of examination employed in many diverse academic regulation, by tradition in the social sciences, but also in market research and further contexts. Qualitative researchers plan to collect an in-depth understanding of people actions and the reasons that manage such behavior. The qualitative method testing the why and how of decision making, not just what, where, when. Thus, minor but focused samples are often used than large amount of samples. Qualitative research analysis procedures create valuable information only on the particular cases studied, and any more general terminations are only suggestion comments. Quantitative research methods can then be used to look for experimental support for such research theories[9].

D. *Expansion of Categories:* The expansion of tweets divided into following 7 categories are:

1. *#Diversity:* Our analyses with this category shows that, engineering student minds diverts from major engineering stream or like thought of switching majors. Also suggest students perceive a significant lack of females in engineering. Major Diversity issue is reflected by the complaints that includes about faculty professors. jobs or campus placements, distraction, confusion, concentration, etc related issue s. The issue here is not lack of diversity, but rather that students have difficulties embracing the diversity, because of many culture conflicts.

2. *#Negative Emotions:* There are a lot of negative, sad like emotions flowing in the tweets. The hash tag #engineering Problems has a negative connotation. We only categorize a tweet as negative emotions when it specifically expresses negative emotions such as sad, anger, sickness, depression, disappointment, boring, fear, hate, hurt, sick, etc.

3. *#Workload:* Our analyses with this category show that, classes, homework, exams, and labs dominate the students life. Students express a very stressful experience in engineering. This category contains tweets related to Stress, load of work, tension with engineering life and tough to do.

4. *#Tough:* 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 issue s that engineering students have but seen in very small volumes. Examples of these issue s include curriculum problems, lack of motivation, procrastination, career and future worries, identity crisis, and physical health problems.

5. *#Stress:* Students are mostly stressed with schoolwork. It is necessary for students to get help with how to manage stress and get emotional support.

6. *#Lack Of Sleep:* Our analyses with this category find that, less sleep problems are widely common among engineering students. Students frequently suffer from less sleep and nightmares due to heavy study load and stress. Chronic less sleep or low-quality sleep can result in many psychological and physical health problems, hence this issue needs to be addressed.

7. *#Missing Social Life:* The analyses with this category show that, students need to sacrifice the time for social engagement in manner to do homework, and to prepare for classes, studies and exams. Many tweets mention having to sacrifice the time enjoying holidays and special occasions with family and friends. Social engagement can provide support for releasing stress and is beneficial for learning. Each category reveals one problem or difficulty that engineering students have in their educational life. It establishes that number of tweets fit in to more than one category. For example, "I am fed up of study why I am not in other school? Hate myself being in engineering. Too much stuff and way too difficult. Nothing enjoy is there" comes into Stressful, Negativity and Major Diversion too at the same time. Hence one tweet can have many categories. This is a multi-label classification as contrast to a single label classification in which each tweet can fall only in one category. The number of categories where one tweet fit in to are called tweets labels otherwise label set.

## E. Mining with Tweets Classification

There are some popular classification algorithms that include Naive Bayes, Support Vector Machine (SVM), Logistic Regression, Maximum Entropy, Decision Tree, and Boosting. So Based on the number of classes which 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 models. Single-label classification means each data point can only fall into one single class where all classes are mutually exclusive. Multi-label classification, however, allows each data point to fall into several or multiple classes at the same time.

Most existing studies on tweet classification are either binary classification on relevant and irrelevant content, or multi-class classification on generic classes such as news, events, opinions, deals, and private messages.

Sentiment analysis is another very popular three-class classification on positive, negative, or neutral emotions/opinions. Sentiment analysis is very useful for mining customer opinions on products or companies through their reviews or online posts.

However, in 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 understanding of students' experiences especially their learning-related issue s and problems. Even for a human interpretation to determine what student problems a tweet indicates is a more difficult task than to determine the sentiment of a tweet. Hence, our study needs a qualitative analysis, and is impossible to do in a fully unsupervised way. Therefore, sentiment analysis is not applicable to our study. We implemented a multi-label classification model where we allowed one tweet to fall into several categories at the same time.

## 3. LITERATURE SURVEY

A. *Academic Pathways Study (APS) conducted by the Center for the Advancement of Engineering Education (CAEE):* One of the research projects regarding engineering students educational life experiences is the Academic Pathways Study (APS) conducted by the Center for the Advancement of Engineering Education (CAEE). They used so many research methods that includes surveys, structured interviews; semi-structured interviews, engineering design task, and small focus groups[10].

B. *Educational Data Mining: A Review of the State of the Art:* Educational Data Mining (EDM) is an emerging interdisciplinary research area that contract with the development of methods to explore data generating in an educational context. Educational Data Mining(EDM) uses computational approaches to analyze educational data in a manner to study educational issue d questions. This paper surveys the most relevant studies carried out in this field. First, it introduces EDM that is Educational Data Mining and describes various group of users, types of educational system environments, and the information they provide[6][7].

C. *Enhancing Learning with Visualization Techniques:* Information visualization is a strong means of making sense of this data that has emerged from study in human computer interaction, computer science, graphics, visual design, psychology, and quantitative data analysis. They have been used in drawings and maps for around thousands of years. This analyses how more novel visualization techniques can be used to improve several activities during the learning processes: discovering and understanding educational system resources, collaboration with educators and teachers, reflecting about educators progress, and constructing design of learning experiences[5].

D *Text-based mood classification:* Mood is a powerful form of sentiment expression, analyzing a state of the mind such as being happy, sad or angry. Social media texts are rich in sentiment and this elaborate several fundamental issue s related to mood sensing from these texts and novel applications of this information. Text-based mood classification and clustering, as a sub-problem of opinion and sentiment mining[3].

E. *Social networks mining for analysis and modeling drugs usage:* This paper survey presents approach for mining and analysis of data from social media sites which is based on using Map Reduce model for processing huge amount of data and on using composite applications for performing more sophisticated analysis which are executed on environment for distributed computing based cloud platform. This proposal is for creation characteristics of users who write about drugs and to estimate factors that can be used as part of model for prediction drug usage level in real-world[2].

## 4. METHODOLOGY

A. *Flow of Research:* In this system, there is an human investigative procedure to find the required appropriate data with the help of Twitter hash(#) tags.

Collecting of tweets using the hash tag engineering-Problems (#engineering Problems) and stored it into data storage which corresponding to the step 1 and 2 resp. in Figure 1.

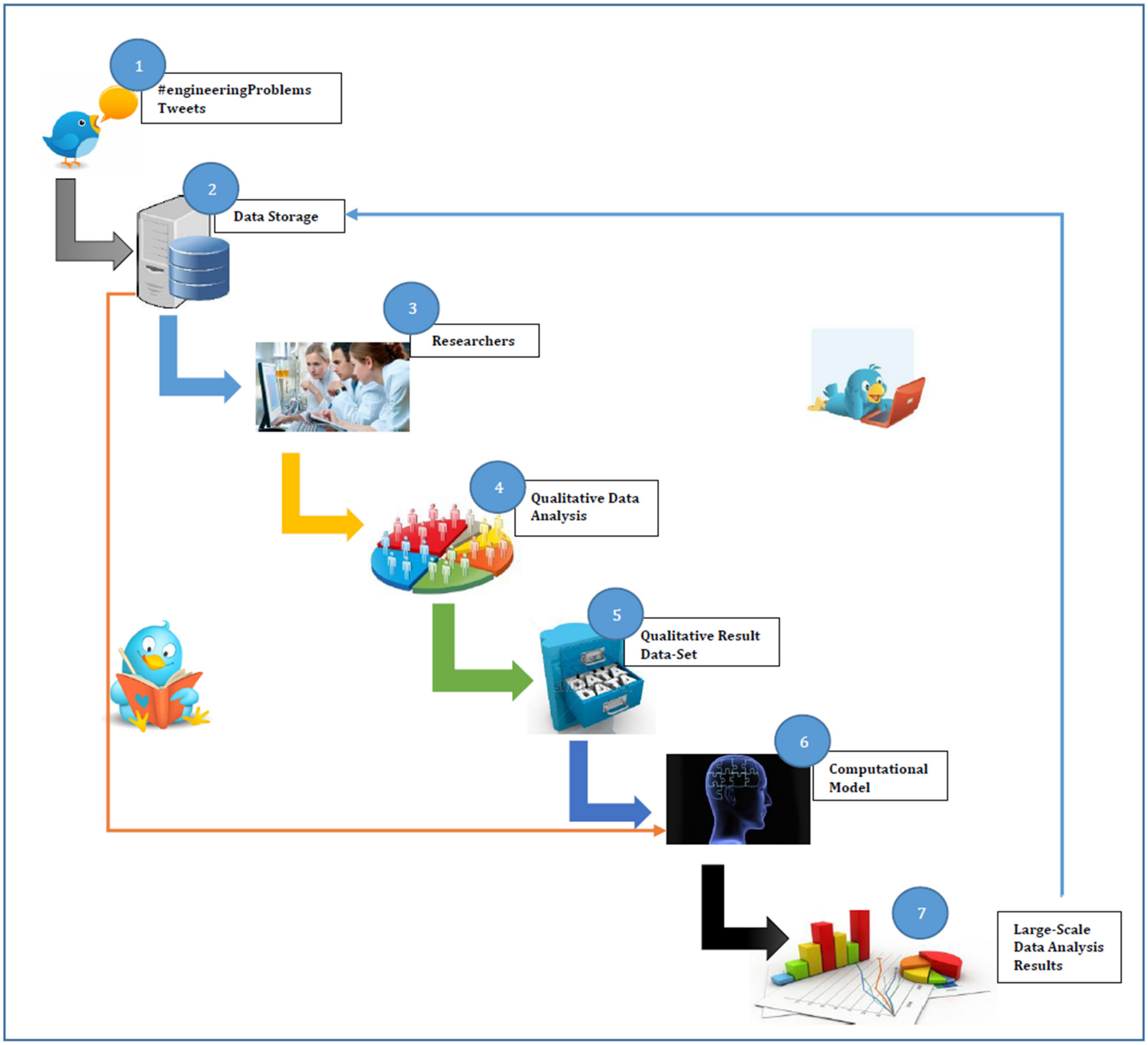


**Figure 1: System Flow Architecture**

Next in the step 3 and 4 of Figure 1. the human interpretation analysis is perform by Researchers on the sample of #engineeringProblems i.e. Qualitative contents analysis.

In Step 5 the outcome of qualitative analysis by human interpretation is stored in another database, it is found that the major issue s that comes into engineering students fall into number of defined categories.

In step 6, Based on these specified categories, a multi-label Naive Bayes classification algorithm is executed for this particular classification. The classification algorithm is applied by system to prepare a Trainer-detector that help to recognize the engineering students problems.

In step 7, based on computational model detector the results are provided The results are provided by step 7 help education researchers to identify at risk students and find out solutions on them and make decisions on proper interference to preserve them to make a good education system.

B. *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, sooo muuchh, and Monnndayyy. Besides, common stopwords 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 #engineeringProblems 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 issue s. So we substituted words ending with nt 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 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.

C. *Naive 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 Naive Bayes classifier to be very effective on our data set compared with other state-of-the-art multi-label classifiers.

The Naive Bayes classifier is a straightforward probabilistic classifier which is based on Bayes theorem with strong and nave self-government assumptions. It is one of the most basic text categorization method with various applications in email spam exposure, private mail sorting, document categorization, language discovery and sentiment discovery.

Naive Bayes executes well in many difficult real-world troubles. Even though it is frequently outperformed by other techniques such as boosted trees, Max Entropy, Support Vector Machines etc., Naive Bayes classifier is extremely efficient since it is less computationally and it requires a small amount of preparation information. One well-liked way to execute multi-label classifier is to convert the multi-label organization problem into multiple single-label categorization problems.

One popular way to implement multi-label classifier is to transform the multi-label classification problem into multiple single-label classification problems. One simple transformation method is called one-versus-all or binary relevance. 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 Naive 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 = $\{w_1, w_2, ..., w_N\}$, and a total number of L categories W = $\{c_1, c_2, ..., c_L\}$. If a word $w_N$ appears in a category $c$ for $m_{w_n c}$ times, and appear in categories other than $c$ for $m_{w_n c}$ times, then based on the maximum likelihood estimation, the probability of this word in a specific category $c$ is,

$$p(w_n | c) = \frac{m_{w_n c}}{\sum_{n=1}^{N} m_{w_n c}}$$

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

$$p(w_n | c') = \frac{m_{w_n c'}}{\sum_{n=1}^{N} m_{w_n c'}}$$

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,

$$p(c) = \frac{C}{M'}$$

and the probability of other categories $c'$ is

$$p(c') = \frac{M - C}{M}$$

## 5. RESULT ANALYSIS

The Figure 2 shows that main user interface of the project. In this interface there are four major sequential parts regards with flow of proposed system includes Gathering Online Tweets, Clustering of tweets according to aspects, Train tweets, and finally Testing tweets for sentiment result analysis.

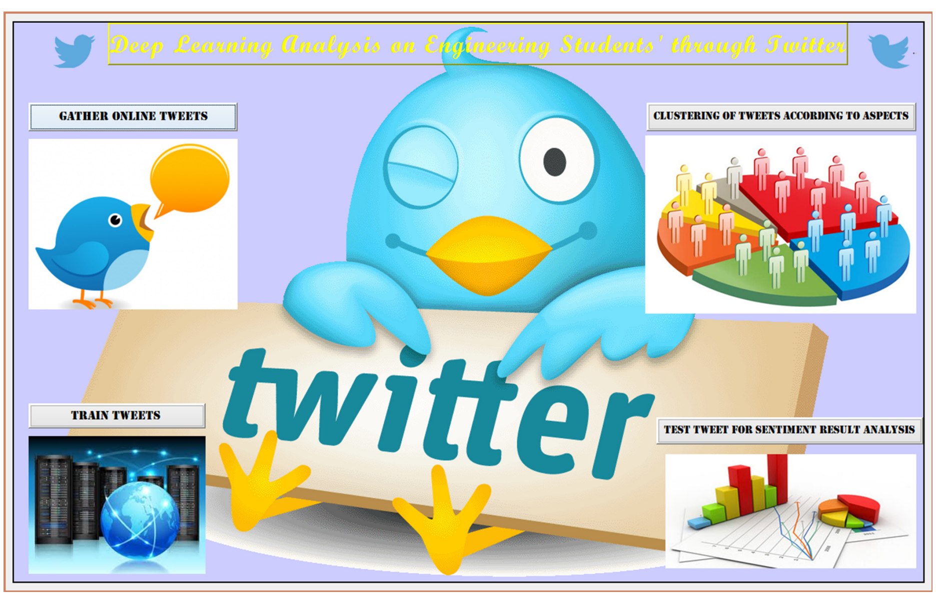


**Figure 2: Main User Interface with four modules**

| Category | Tweet |
|---|---|
| #Diversity | RT @TheTestDoctor: Setting expectations - Scotty, when will the warp drive be fixed? #startrek #engineeringproblems https://t.co/JzyYqGme75 |
| #WorkLoad | It's high time we change the name of Airoli to Awwroli. #EngineeringProblems |
| #Diversity | It's high time we change the name of Airoli to Awwroli. #EngineeringProblems |
| #NegativeEmotions | 2 hours of homework done. Now I get the joy of studying for 3 hours... #engineeringproblems |
| #NegativeEmotions | 2 hours of homework done. Now I get the joy of studying for 3 hours... #engineeringproblems |
| #NegativeEmotions | Exam preparations have me like #sigh #engineeringproblems #sppu https://t.co/xZV1x1PlEq |
| #Diversity | Setting expectations - Scotty, when will the warp drive be fixed? #startrek #engineeringproblems https://t.co/JzyYqGme75 |
| #NegativeEmotions | RT @EChaupal: Select the best option :)\n#Engineering #engineer #engineeringproblems #EngineeringChallenge https://t.co/mNnSlvJAYo |
| #NegativeEmotions | Select the best option :)\n#Engineering #engineer #engineeringproblems #EngineeringChallenge https://t.co/mNnSlvJAYo |
| #Diversity | RT @engineerproblem: When the test crushed the entire class and your professor has no mercy. #engineeringproblems https://t.co/yNO0iA7nw2 |
| #NegativeEmotions | RT @engineerproblem: When the test crushed the entire class and your professor has no mercy. #engineeringproblems https://t.co/yNO0iA7nw2 |
| #WorkLoad | RT @LeeAnneOfNerds: When you spend more time with math than you do with friends (or people for that matter) #EngineeringProblems |
| #MissingSocialLife | RT @LeeAnneOfNerds: When you spend more time with math than you do with friends (or people for that matter) #EngineeringProblems |
| #WorkLoad | When you spend more time with math than you do with friends (or people for that matter) #EngineeringProblems |
| #MissingSocialLife | When you spend more time with math than you do with friends (or people for that matter) #EngineeringProblems |
| #LackOfSleep | *working on a project*\n"Oh look it's 5am.. I better go sleep now so I can wake up early tomorrow" #engineeringproblems |
| #WorkLoad | *working on a project*\n"Oh look it's 5am.. I better go sleep now so I can wake up early tomorrow" #engineeringproblems |
| #WorkLoad | calculated result off by 200%? System dependent scaling factor = 2 #ScienceOnABudget #engineeringproblems |
| #NegativeEmotions | RT @Sudharsan_ak: Retweet if ENGINEERING F***ed you also like this!!!!!!????????\n\n#EngineeringChallenge #engineeringproblems https://t.co/Ottq6 |
| #Diversity | Things in motion tend to stay in motion, while things at rest tend to stay at rest.\nThis is why i must not sleep. haha #engineeringproblems |
| #LackOfSleep | Things in motion tend to stay in motion, while things at rest tend to stay at rest.\nThis is why i must not sleep. haha #engineeringproblems |
| #Diversity | Things in motion tend to stay in motion, while things at rest tend to stay at rest.\nThis is why i must not sleep. haha #engineeringproblems |
| #NegativeEmotions | Things in motion tend to stay in motion, while things at rest tend to stay at rest.\nThis is why i must not sleep. haha #engineeringproblems |
| #NegativeEmotions | Retweet if ENGINEERING F***ed you also like this!!!!!!????????\n\n#EngineeringChallenge #engineeringproblems https://t.co/Ottq6fcnkL |
| #LackOfSleep | Update: the eye twitch continues, despite sleep AND caffeine. What could be the cause? #ElectionResults #engineeringproblems #examstress |

**Figure 3: Categorized gathered tweets**

The Figure 3 shows that the tweets after clustering according to aspects like #Diversity, #LackOfSleep, #NegativeEmotions, #Workload, #Tough, #Stress, #MissingSocialLife, etc. by using qualitative analysis and some machine learning techniques.

| Category | tweet | processedtweet |
|---|---|---|
| #NegativeEmotions | RT @engineerproblem: -How'd midterms go?\n-Oh not ... | midterms go?\n-oh bad. too. walk crying*\n |
| #MissingSocialLife | Doctor : do you drink a lot?\nMe : how many is a l... | doctor drink lot?\nme lot? |
| #NegativeEmotions | Engineering colleges are like a shoppers stops eve... | engineering colleges shoppers stops overrated una... |
| #LackOfSleep | Doctor : do you drink a lot?\nMe : how many is a l... | doctor drink lot?\nme lot? |
| #Diversity | Engineers, Designers, Draftsman, don't miss this. ... | engineers, designers, draftsman, this. engineerin... |
| #Diversity | One more placement and one more failure..\nThe sag... | placement failure..\nthe saga continues engineeri... |
| #NegativeEmotions | betakyakarrhehophobia : afraid to be asked by rela... | betakyakarrhehophobia afraid asked relatives pado... |
| #NegativeEmotions | I'm never sure if I hate NX or Solidworks more. I ... | hate nx solidworks more. guess depends assignment |
| #Diversity | betakyakarrhehophobia : afraid to be asked by rela... | betakyakarrhehophobia afraid asked relatives pado... |
| #Diversity | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |
| #NegativeEmotions | RT @engineerproblem: -How'd midterms go?\n-Oh not ... | midterms go?\n-oh bad. too. walk crying*\n |
| #Diversity | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |
| #NegativeEmotions | Engineering colleges are like a shoppers stops eve... | engineering colleges shoppers stops overrated una... |
| #Diversity | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |
| #Diversity | Engineers, Designers, Draftsman, don't miss this. ... | engineers, designers, draftsman, this. engineerin... |
| #Diversity | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |
| #NegativeEmotions | betakyakarrhehophobia : afraid to be asked by rela... | betakyakarrhehophobia afraid asked relatives pado... |
| #WorkLoad | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |
| #Diversity | betakyakarrhehophobia : afraid to be asked by rela... | betakyakarrhehophobia afraid asked relatives pado... |
| #NegativeEmotions | @notmegangreen I feel ya #engineeringproblems | feel ya |
| #Diversity | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |
| #Stress | RT @HugotInhinyero: I think we can all relate to t... | think relate these. |
| #Diversity | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |
| #WorkLoad | RT @HugotInhinyero: I think we can all relate to t... | think relate these. |
| #Diversity | @Ae_Kaun the waking up at 5,6 happens at the day b... | waking day exams sleeping study holidays. gtu |

**Figure 4: Text Pre-Processed Tweets**

The Figure 4 shows the result after applying text pre-processing steps which is explained in section IV(B). Thus, after pre-processed the tweets it gives some valuable keywords that shown in third column in Figure 4, it will definitely helpful for further process.

| Category | Keyword | Count_in_C | Count_in_other_than_C |
|---|---|---|---|
| NegativeEmotions | shit | 2 | 4 |
| NegativeEmotions | assignment | 6 | 12 |
| Stress | think | 26 | 36 |
| Stress | relate | 8 | 8 |
| Stress | these. | 8 | 8 |
| Diversity | clean | 2 | 0 |
| Diversity | mechanical | 2 | 4 |
| Diversity | engineering | 60 | 122 |
| Diversity | books | 2 | 2 |
| WorkLoad | think | 28 | 34 |
| WorkLoad | relate | 8 | 8 |
| WorkLoad | these. | 8 | 8 |
| WorkLoad | lost | 4 | 4 |
| WorkLoad | followers | 2 | 0 |
| WorkLoad | started | 4 | 16 |
| WorkLoad | jet | 2 | 0 |
| WorkLoad | pack | 2 | 0 |
| WorkLoad | stuff. | 2 | 0 |
| WorkLoad | stopping | 2 | 0 |
| WorkLoad | tho | 2 | 0 |
| WorkLoad | gottablast | 2 | 0 |
| WorkLoad | jetfuel | 2 | 0 |
| NegativeEmotions | sunday | 2 | 0 |
| NegativeEmotions | wasted | 2 | 0 |
| NegativeEmotions | day | 6 | 18 |

**Figure 5: Count of words in respective category & other categories too**

The Figure 5 shows the result analysis of finding the count of every word which is came after text preprocessing with respect to respective category and also with other categories as shown in column 3rd and 4th resp. in Figure 5.

| Category | keyword | Probability_in_c | Probability_in_other_than_c |
|---|---|---|---|
| NegativeEmotions | fear | 0.00184729 | 0.00123077 |
| NegativeEmotions | love | 0.00432099 | 0.00123077 |
| NegativeEmotions | engineering. | 0.00123077 | 0 |
| Stress | think | 0.0449827 | 0.0633803 |
| Stress | relate | 0.0134228 | 0.0134228 |
| Stress | these. | 0.0134228 | 0.0134228 |
| WorkLoad | think | 0.0150376 | 0.018319 |
| WorkLoad | relate | 0.0042508 | 0.0042508 |
| WorkLoad | these. | 0.0042508 | 0.0042508 |
| NegativeEmotions | sunday | 0.000615006 | 0 |
| NegativeEmotions | wasted | 0.000615006 | 0 |
| NegativeEmotions | day | 0.00184729 | 0.00556242 |
| NegativeEmotions | sleeping | 0.00184729 | 0.00370142 |
| NegativeEmotions | suddenly | 0.00432099 | 0 |
| NegativeEmotions | mid-night | 0.000615006 | 0 |
| NegativeEmotions | realize | 0.00246457 | 0.00246457 |
| NegativeEmotions | shit | 0.000615006 | 0.00123077 |
| NegativeEmotions | assignment | 0.00184729 | 0.00370142 |
| Diversity | clean | 0.000877963 | 0 |
| Diversity | mechanical | 0.000877963 | 0.00175747 |
| Diversity | engineering | 0.027027 | 0.0565338 |
| Diversity | books | 0.000877963 | 0.000877963 |
| WorkLoad | lost | 0.00212089 | 0.00212089 |
| WorkLoad | followers | 0.00105932 | 0 |
| WorkLoad | started | 0.00212089 | 0.00853789 |

**Figure 6. Probability of words in respective category & other categories too**

The Figure 6 shows the result analysis of finding the probability of every word which is done after find the count of words with respect to respective category and also with other categories as shown in column 3rd and 4th resp. in Figure 6.

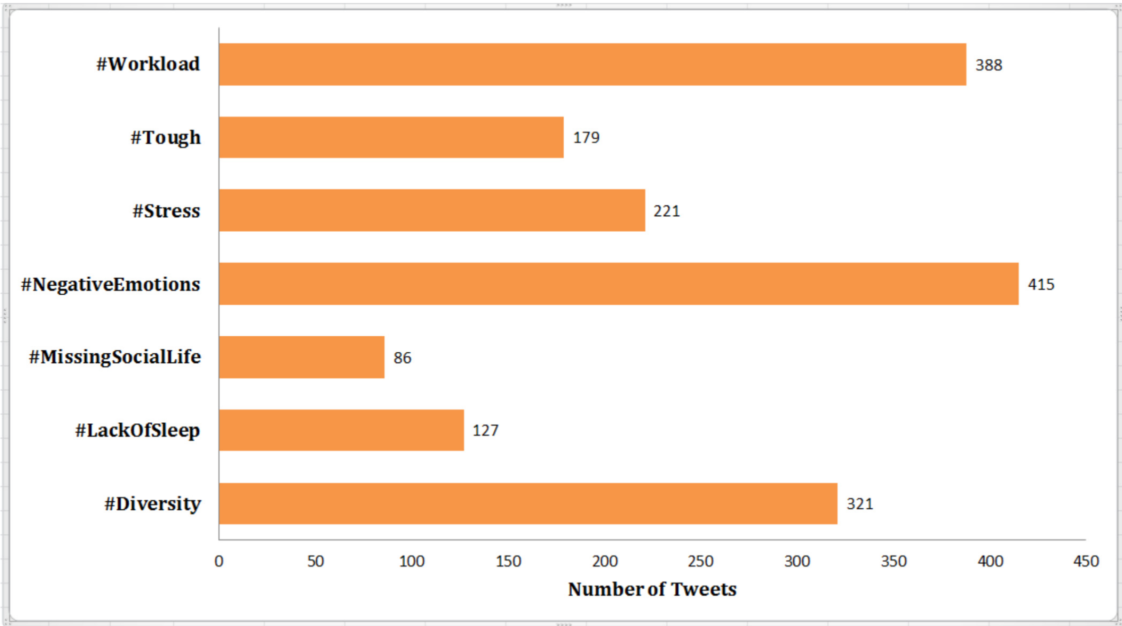


**Figure 7: Graphical representation shows number of tweets according to aspect**

The Figure 7 shows the final result with the help of visualization technique that displays number of tweets according to respective categorized aspects with more accuracy in the format of graphical representation (chart).

## 6. LIMITATIONS & FUTURE WORK

There are a number of limitations, which also lead to many possible directions for future work:

Not all students are active on Twitter, so may only find the ones who are more active and more likely to expose their thoughts and feelings.

The fact that the most relevant data find on engineering students learning experiences involve complaints, issue s, and problems does not mean there is no positive aspects in students learning experiences. Future work can compare both the good and bad things to investigate the tradeoffs with which students struggle.

Here I'm going to consider the prominent themes with relatively large number of tweets in the data. There are a variety of other issue s hidden in the long tail. Several of these issue s may be of great interest to education researchers and practitioners.

## 7. CONCLUSION

Mining social media data is helpful to educational researchers in learning analytics, educational data removal, and also learning skills. It gives a proper way to observing and studying social medium statistics that combine the main restrictions of both human physical qualitative analysis and large scale computational study of user produced textual contents. It mainly notifies educational manager, and other applicable education assessment makers to expand further accepting of engineering students institution understanding.

## Acknowledgement

I would like to express my gratitude to all those who gave me the possibility to complete this work of research study. I want to thank the our Department to do the necessary research work and to use departmental data. I am deeply indebted to my Project Guide Prof. U. B. Chavan, whose help, stimulating suggestions and encouragement helped me in all the time of writing of this paper. I am furthermore thankful to Head of Department Dr. S. P. Sonavane, who gave and confirmed this permission and encouraged me to go ahead with my work of study. My all colleagues supported me in my study work. I want to thank them for all their help, support, interest and valuable hints.

## REFERENCES

[1] X. Chen, M. Vorvoreanu, and K. P. C. Madhavan, "Mining social media data for understanding students' learning experiences," IEEE Transactions on Learning Technologies, Vol. 7, No. 3, pp. 246–259, 2014.

[2] A. Yakushev and S. Mityagin, "Social networks mining for analysis and modeling drugs usage," Procedia Computer Science, Vol. 29, pp. 2462–2471, 2014.

[3] T. Nguyen, D. Phung, B. Adams, and S. Venkatesh, "Mood sensing from social media texts and its applications," Knowledge and Information Systems, Vol. 39, No. 3, pp. 667–702, 2014.

[4] E. Popescu, "Providing collaborative learning support with social media in an integrated environment," World Wide Web, Vol. 17, No. 2, pp. 199– 212, 2014.

[5] J. Klerkx, K. Verbert, and E. Duval, "Enhancing Learning with Visualization Techniques," Handbook of Research on Educational Communications and Technology, No. 1999, pp. 791–807, 2014.

[6] R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009 : A Review and Future Visions," Journal of Educational Data Mining, Vol. 1, No. 1, pp. 3–16, 2009.

[7] C. Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Vol. 40, No. 6, pp. 601–618, 2010.

[8] C. J. Atman, S. D. Sheppard, J. Turns, R. S. Adams, L. N. Fleming, R. Stevens, R. A. Streveler, K. A. Smith, R. L. Miller, L. J. Leifer, K. Yasuhara, and D. Lund, "Enabling Engineering Student Success: The Final Report for the Center for the Advancement of Engineering Education," Center for the Advancement of Engineering Education, p. 224, 2010.

[9] J. Tang, Y. Zhang, J. Sun, and J. Rao, "Quantitative study of individual emotional states in social networks," Affective Computing, . . ., Vol. 3, No. 2, pp. 132–144, 2012.

[10] M. Clark, S. D. Sheppard, C. J. Atman, L. N. Fleming, and R. L. Miller, "Academic Pathways Study : Processes and Realities," Chart, 2008.