

# Forum Based Data Mining Tool for Teaching Learning Experience

Govindraj B. Chittapur<sup>1</sup> , Veena B. Desai<sup>2</sup>

<sup>1</sup>Dept of Computer Applications Basaveshwar Engineering College Bagalkot India;

<sup>2</sup>Dept of Computer Applications Basaveshwar Engineering College Bagalkot India;

## Abstract

*Today we are living in a social media world, we use to give much information to the social forum, by referring as twitter information as social forum and identifying teaching learning experience through given set of information. Students' informal conversations on social media shed light into their educational experiences opinions, feelings, and concerns about the learning process. Data from such un-instrumented environments can provide valuable knowledge to inform student learning. We used Navie Bayes multi classifier algorithm used for classify tweet based student opinion about teaching learning experience even by using Maximum likelihood method is used to computer and compare classifier result.*

**Keywords:** Data Mining, Forum, Navie-Bayes, Maximum-Likelihood, Stop words, Student problems, Probability computation,

## 1. Introduction

Data mining is an interdisciplinary subfield of computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems.

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. From the raw analysis step, it involves database and data management aspects, data preprocessing model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization and online updating.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records, unusual records and dependencies. These patterns can then be seen as a kind of summary of the input data and may be used in further analysis. The KDD process includes data collection, data preprocessing, data interpretation and opinion mining.

Opinion mining is a natural language processing for tracking the mood of the public sentiment. This sentiment mining is the investigation that looking for the total count of Likes, Shares or Comments that we get, based on those comments we identify the problem definition on teaching

learning experiences related to forum. Social media sites such as face book, linked in and twitter provides social and professional interaction support. On various social media sites, students discuss and share their opinion and emotional feeling everyday encounters in an informal 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. The abundance of social media like twitter data provides opportunities to understand students' experiences. These are all students' experiences becomes an input for developing this project.

## 2. Literature Survey

Many research and development work has been incorporated in the area of data mining related with sentiment analysis, here we consider for related with data mining tool for teaching learning experience as available solution and their features. In the following table we discuss some of the investigated methods and research algorithm discussed.

**Table 2.1** Review of Literature survey

Sl.No	Author and Publications	Research Paper Title	Proposed Work
1.	Xin Chen, Student Member, IEEE, Mihaela Vorvoreanu, and Krishna Madhavan IEEE Transaction On Learning Technologies, VOL. 7, NO. 3, JULY-SEPTEMBER 2014	Mining Social Media Data for Understanding Students' Learning Experiences	NAEIVE BAYES classifier based student problem of College of Engineering, Purdue University
2.	L.R.Jeevitha , R.Priyanka, P.Deepa, A.N.Sasikumar International Journal of Advanced Research in Computer Science Engineering and	Mining Social Media for Understanding Students' Learning Experiences	Data mining preprocessing algorithms and Data cleaning by multi label classification model

	Information Technology Volume: 4 Issue: 3 23-May- 2015.ISSN_NO: 2321-3337		
3.	Pallavi K. Pagare International Journal of Computer Applications (0975 – 8887) Innovations and Trends in Computer and Communication Engineering (ITCCE-2014)	Analyzing Social Media Data for Understanding Student’s Problem	Text Pre- processing algorithm
4.	Tina R. Patil, Mrs. S. S. Shrekar Journal of sci.Education, Vol.86, No.1, Pp.7-15, 2000	Performance Analysis Of Naive Bayes And J48 Classification Algorithm For Data Classification	Introduces an algorithm for learning from labeled and unlabeled documents based on the combination of Expectation- Maximization (EM) and a naive Bayes classifier.
5.	Bo Pang and Lillian Lee e Morgan & Clay Pool Publishers, Pp. 54-58, 2008.	A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based On Minimum Cuts	relation between subjectivity detection and polarity classification, showing that subjectivity detection can compress reviews into much shorter extracts that still retain polarity information at a level comparable
6.	Kamal Nigam, Andrew Kachites Mccallum, Sebastian Thrum, Tom Mitchell, Machine Learning, 39, Kluwer Academic Publishers. Printed In The Netherlands, Pp. 103–134, 2000	Text Classification From Labeled And Unlabeled Documents Using EM”	Text classifiers can be improved by augmenting a small number of labeled training documents with a large pool of unlabeled documents.

### 3. Design Issues of forum based Teaching Learning Experience

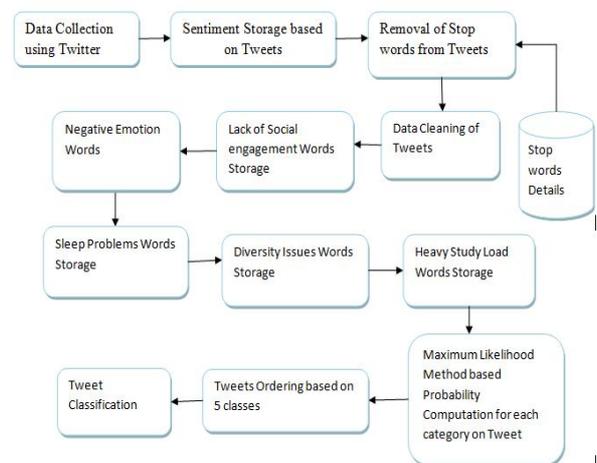
By considering of the literature survey the work basically focused on static data, here we are proposing dynamic approach for student learning experience by considering

the twitter for collecting student experience with different forum.

We chose to focus on engineering students’ posts on Twitter about problems in their educational experiences mainly because:

1. All most all Engineering colleges have long been struggling with student recruitment and maintenance issues. Engineering graduates constitute a significant part of the future workforce and have a direct impact on the economic growth and global competency.
2. Based on understanding the issues and problems of students, representative and instructors can make more informed decisions on proper interventions and services that can help the students to overcome barriers in learning.
3. Twitter is the most popular social media site. Its content is mostly public and very concise that means there won’t be more than 140 characters per single tweet. Twitter provides free APIs that can be used to stream data. Therefore, it’s better to start analyzing students’ emotions or feelings from the posts on Twitter.

By considering data set from twitter here we propose Navie Bayes based classifier to classify and recognize categories based on maximum likelihood dynamic multi classifier to identify student problems.



**Fig 3.1** Proposed architectural diagram of Forum Based Data Mining Tool for Teaching Learning Experience

#### 3.1.1 Data Collection using Tweet

Here in this data collection we collect the tweet data which are taken from twitter interface among the set of accounts. This twitter data retrieval is done by using twitter API. This Oath API is used to authenticate the open source framework with the twitter application. Here we have tendency to collect student information for their suggestions that need to be followed for the tweet information.

### 3.1.2 Sentiment Storage based on Tweets

In this sentiment storage, mainly sentiment tweets are based on the twitter interfaces that are collected from the students or user. And these tweets are the process of storing the data about the tweets into the relational storage in terms of *TwitterId*, *TwitterDesc* and *UserId*. Where Twitter Id describes the unique Id which is associated with the tweet, TwitterDesc describes the actual description of tweet and UserId describes the specific user Id which is associated specific user.

### 3.1.3 Stop Words

These are the set of words which do not have any specific meaning. The data mining has some set of standard stop words. These stop words are filtered out before or after processing of natural language data. There is no definite list of stop words in which all the tools use and such a filter is not always used. In this project we mainly consider these stop words for the work to be executed.

The various 5 categories used in the algorithm implementation will be associated with a set of tokens as described below

Category Name	Category Words
Heavy Study Load	hour, homework, exam, day, class, work, divine, problem, study, week, too much, 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, shit, exam, negtoken, week, class, hell, engineer, suck, study, hour, homework, time, equate, FML, lab, sad, bad, day, feel, tire, damn, death, hard
Sleep Problems	sleep, hour, night, negtoken, bed, all-night, exam, homework, nap, coffee, time, study, more, work, class, dream, lady engineer, 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, too much, walk, people, teach, understand, chick, China, foreign, out, white, black

## 4. Proposed Novel Approaches for Development Process

### 4.1 Novel Classifier for modified Navie Bayes Algorithm:

We propose a modified Navie Bayes algorithm for development of identifying and categorizing student problems through social forum. In this development process we collect the hash tags which are authenticated through OAuth API. Through this authenticated hash tags we collect the tweets from the twitter interface. Then among collected tweets we perform the computation of specific category on each tweet. First we compute the probability value of all the tweets, then contingency and enhanced contingency is calculated based on the probability value further based on the computed value we classify the tweet on the bases of category then we

compute the classifier the count of tweet. And shown in the below diagrammatical representations.

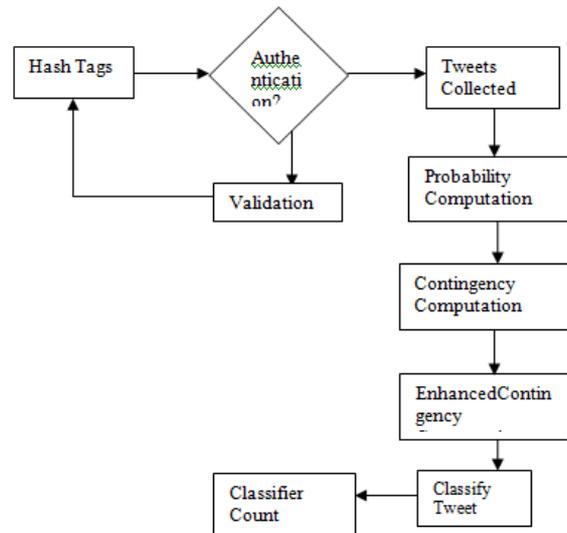


Figure 3.3 Development process to classify the count using tweet

### 4.2 Maximum Likelihood Estimation Algorithm:

This algorithm explains the maximum likelihood estimation algorithm used to classify the categories to identify the categories of student problem by referring tweet API.

Let us we take S is some tweet description

**Step 1:** Define categories  $c = \{c_1, c_2, c_3, \dots, c_n\}$

**Step 2:** Read data from a database.

**Step 3:** Divide S into sub works  $\{w_1, w_2, w_3 \dots w_n\}$  split.

**Step 4:** Check sub words  $\{w_1, w_2, w_3 \dots w_n\}$  for every categories

**Step 5:** if words match with categories  $\{c_1, c_2, c_3 \dots c_n\}$  increment the counter for that categories Else put that in "other" categories.

**Step 6:** Find probability of each category.

Where C belongs to different specified categories and W is work  $\{w_1, w_2, w_3, w_4 \dots w_n\}$  and sub work of tweet sentence

## 5. Result and Discussion

We collect the data from twitter interface and develop the process through java programming environment, For computation approach we use open source environment. Here we showed the result with each module and evaluation.

HashTags	
HASH TAG ID	HASHTAG
1	studentissues
2	studentclass
3	studentreads
4	studentrajesh
5	studentdevi
6	students project
7	students talk
8	students demo
9	studentry

5.1 Collected Hash Tag Description of OAuth Interface.

This outcome of the result shows the hash tags that are added to the twitter data and here we add the value associated with standard key word #student. Other than the value not associated with #student key words are not valid value.

Tweet ID	Tweet Details	User ID	Screen Name	Language	Hash Tag
168	@Baraboy When you score 249 in jambi and cut off mark is 250 @sachinringga @e...	Toluwinmi	_trouzi_	en	studentclass
169	RT @SDP0521976: Kurang ngomong bkl krt? Friday time profes...	divi suabro	divisuarndi	in	studentclass
170	RT @Alye_Shinchi: Lagi basic banget untuk megeja jbra klabenan serent...	Sieran StudentClass	StudentClass_76	in	studentclass
171	@SDP0521976: Weh smlc bak kash aqua hata hell coffee dit zomuch asa...	Sieran StudentClass	StudentClass_76	in	studentclass
172	@beraniboo: @StudentClass_76 Mas baik mas out bed shi exam class...	South Sumatera	SDP0521976	in	studentclass
173	@Minnid4AP: @beraniboo Mas out mas hard exam study engineer equate...	Sieran StudentClass	StudentClass_76	in	studentclass
174	@StudentClass_76: well homework awake long beautiful bed...	Aku Wlqgy	Alye_Shinchi	en	studentclass
175	@beraniboo: @StudentClass_76 Mas good mas today library nrostep rush beautiful...	Aek	Minnid4AP	en	studentclass
176	@StudentClass_76: Mas culuk mas shi long homework engineer skill...	Samp'Udaibaa	beraniboo	in	studentclass
177	@StudentClass_76: Kurang ngomong bkl krt? die spend week engineer regatkan...	SEPAH PANG JOMBANG	SDP0521976	in	studentclass
178	RT @LUD0521976: Pendaftaran sudah di buka, untuk vanto trun seblan sblkan...	Sieran StudentClass	StudentClass_76	in	studentclass
179	RT @Klarnet123: bklah WorkPlace 100 adede - #100TAVOZ.NET https://t.co/0v...	e_lj_lbr	no	no	students project

5.2 Description of retrieved tweets from the tweet interface.

This outcome of the result shows the tweet information where tweet users have tweeted for the different hash tag.

Stop Word ID	Stop Word
152	latter
153	latterly
154	least
155	less
156	ltd
157	made
159	may
160	me
161	meanwhile
162	might
163	mill
164	mine

5.3 Collected stop word description of standard data mining stop words along with user defining stop words.

This outcome of the result shows stop words with standard stop words and here we can also add user stop words.

Tweet ID	Tweet Details	User ID	Screen Name	Language	Hash Tag
1	rt studentclass's summer sorted studentclass https://w... in qm u ym people equate hour...	Wendee Lembre	1181e5jupnd	en	studentclass
2	rt studentclass un making on life ending wondering w contacting sho thirtheeproblems...	Paktor Sumett	Paktor8500262	en	studentclass
3	baraboy score jambi cut mark sashchinringga studentclass vshimgieia w...	Toluwinmi	_trouzi_	en	studentclass
4	rt schupas studentclass kurang ngomong bkl krt friday time profesor zomuch...	divi suabro	divisuarndi	in	studentclass
5	rt alye shinchi lagi basic banget untuk megeja jbra klabenan serentata le...	Sieran StudentClass	StudentClass_76	in	studentclass
6	schupas weh smlc bak kash aqua hata hell coffee dit zomuch asa...	Sieran StudentClass	StudentClass_76	in	studentclass
7	beraniboo studentclass mas baik mas bed shi exam class...	South Sumatera	SDP0521976	in	studentclass
8	minnid4ap beraniboo mas out mas hard exam study engineer equate...	Sieran StudentClass	StudentClass_76	in	studentclass
9	studentclass homework awake long beautiful bed...	Aku Wlqgy	Alye_Shinchi	en	studentclass
10	beraniboo studentclass mas good mas today library nrostep rush beautiful...	Aek	Minnid4AP	en	studentclass
11	studentclass mas culuk mas shi long homework engineer...	Samp'Udaibaa	beraniboo	in	studentclass
12	studentclass kurang ngomong bkl krt spend week engineer regatkan...	SEPAH PANG JOMBANG	SDP0521976	in	studentclass

5.4 Data cleaning using stop words.

This outcome of the result shows the cleaned tweets from tweets information where stop words are removed from tweeted data.

Category Word	Category	
126	white	Diversity Issues
128	vodka	Heavy Study Load
129	adulthood	Sleep Problems
130	oralexam	Negative Emotion
131	distracted	Heavy Study Load
135	assignment	Heavy Study Load
136	new work	Heavy Study Load
137	misunderstanding	Lack of Social Engagement
138	negative thinking	Negative Emotion
139	late sleeping	Sleep Problems
140	mind diverting	Diversity Issues
141	more time working	Heavy Study Load

5.5 Redefining of category words.

This outcome of the result shows the category word with category.

Tweet ID	Category Name	Count	Total Words	Probability	Negative Probability
152	Diversity Issues	0	23	0	1
153	Heavy Study Load	1	17	0.0588235294117647	0.941176470588235
153	Lack of Social Engagement	1	17	0.0588235294117647	0.941176470588235
153	Negative Emotion	0	17	0	1
153	Sleep Problems	0	17	0	1
153	Diversity Issues	0	17	0	1
154	Heavy Study Load	1	20	0.05	0.95
154	Lack of Social Engagement	0	20	0	1
154	Negative Emotion	0	20	0	1
154	Sleep Problems	0	20	0	1
154	Diversity Issues	0	20	0	1
155	Heavy Study Load	1	18	0.0555555555555556	0.944444444444444

5.6 Probability computation for each category word.

This outcome of the result shows the probability of the tweet among group of tweets from the maximum likelihood method.

Tweet ID	Cat Name
152	Lack of Social Engagement
153	Heavy Study Load
153	Lack of Social Engagement
153	Heavy Study Load
153	Lack of Social Engagement
154	Heavy Study Load
154	Heavy Study Load
155	Heavy Study Load
155	Heavy Study Load
156	Heavy Study Load
156	Sleep Problems
156	Negative Emotion

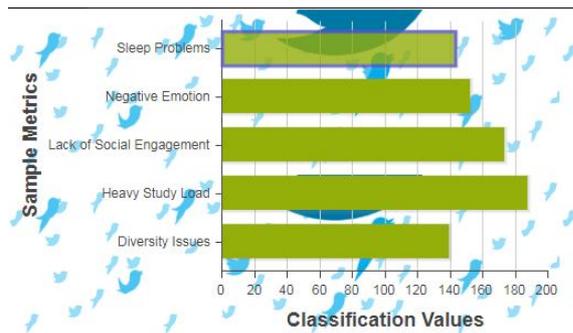
5.7 Classifying the category words.

This outcome of the result shows to which of the category the specified tweet belongs to and here different tweets can belong to the same category.

Count	Cat Name
138	Diversity Issues
186	Heavy Study Load
172	Lack of Social Engagement
151	Negative Emotion
142	Sleep Problems

5.8 Result of classifier count.

This outcome of the result shows the total number of tweets along with category name.



### 5.9 Comparison of various category with classification values.

This outcome of the result shows the final classifier graph with classification values along with simple metrics. It changes whenever tweet data is re-retrieved.

## 6. CONCLUSION

In today's world lot of focus is on the study habits and study processes which improve the knowledge of the students which are really good. But there are no approaches which mainly concentrate on the student strengths. In this proposed research paper we have used twitter interface to retrieve the tweet data and here we have mentioned standard hash tag i.e. student and from here we get the all the user information who have twitted for this particular hash tag we have used. It gives a way to examining social standard information that conquers the main restrictions of both physical qualitative analysis and huge scale computational study of user produced textual content.

Here in this research paper we have used an algorithm that is *Naïve Bayes Multi-Label Classifier* which is implemented on the bases of several steps like Data Collection from twitter, Cleaning the data by removing stop words, removal of non letter and punctuation marks, probability of the words for various categories namely Heavy Study Load, Sleep Problems, Lack of Social Engagement, Negative Emotion and Diversity Issues is estimated.

In this research paper the twitter interface we need to find the accuracy, probability, contingency, enhanced contingency, total count and classifier count for all the specified tweets. Then based on these computed value classical graph is performed as category versus classifier count value. Therefore here we can conclude on average how many students have various categories of problems as well as extend this to the problems faced by which user.

## REFERENCE

[1]. Xin Chen, Student Member, IEEE, Mihaela Vorvoreanu, and Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences" IEEE Transaction On Learning Technologies, VOL. 7, NO. 3, JULY-SEPTEMBER 2014

[2]. L.R.Jeevitha, R.Priyanka, P.Deepa, A.N.Sasikumar "Mining Social Media for Understanding Students' Learning Experiences" International Journal of Advanced Research in Computer Science Engineering and Information Technology Volume: 4 Issue: 3 23-May-2015,ISSN\_NO: 2321-3337

[3]. Pallavi K. Pagare, "Analyzing Social Media Data for Understanding Student's Problem" International Journal of Computer Applications (0975 – 8887) Innovations and Trends in Computer and Communication Engineering (ITCCE-2014)

[4]. Tina R. Patil, Mrs. S. S. Sherekar, "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification" Journal of science .Education, Vol.86, No.1, Pp.7-15, 2000

[5]. Bo Pang and Lillian Lee e, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based On Minimum Cuts" Morgan & Clay Pool Publishers, Pp. 54-58, 2008.

[6]. Ms. Pranjali S. Jadhav, 2Dr. Shirish S. Sane," Understanding Students' Learning Experience By Data Mining Of Social Media".

## AUTHOR



**Veena Desai** received the B. Sc degree in computer Science under Basaveshwar Science College Rani Channamma University during the year 2011 and 2014 and completed MCA Viva Voce Examination 2017.



**Govindraj B. Chittapur** received the M.Sc Technology by research degrees in Computer Science and Technology from PES Research Foundation, PES college of Engineering and Technology, University of Mysore in 2011 and Master of Computer Applications from Shri. Jayachamarajendra College of Engineering, Visheswaraya Technology University in 2005, respectively. He is serving as Editorial Member and Reviewer of International Journal and conference. He has published 17 International Journal and 2 national conferences. He is a professional member of International Economic Research Centre (IEDRC) Hong Kong –China and ACM Associates Computer Science Teachers Associations.