

Classification of Student's E-Learning Experiences' in Social Media via Text Mining

Ms. Priyanka Patel¹, Ms. Khushali Mistry²

¹(Department of CSE, PIET, Vadodara, India, ²(Department of CSE, PIET, Vadodara, India,

Abstract : In today's world, social media is used every individual for expressing their feelings, opinion, experiences' and emotions. Applying data mining on all these emotions expressed in posts, comments and likes called as social media data. In Existing system, only the prominent themes are identified with relatively large number of tweets in the data. There are a variety of other issues hidden in the "others" theme. Several of these issues may be of great interest to education researchers and practitioners. The fact that the most relevant data which are found on engineering students' learning experiences involve complaints, issues, and problems does not mean there are no positive sides in students' learning experiences'. This may entail that social media serve as a good venue for students to utter negative emotions and seek social support. In proposed work, New Label Good Things introduce and using the probability, the common keywords are considered for this label. Doing this, the proposed system is identifying and classifying the e-learning problems faced by student to improve their education quality with respect to their good and positive comments. Naive Bayes multilabel classifier is used for classification of experiences' by finding the probability of words in tweet for each category probability of each label contains how many users. Finally the tweets with new Label will be compared with the rest of the tweets with existing Labels.

Keywords – Classification, social media, multilabel, learning experiences', text mining .

I. Introduction

In today's world we mostly use websites for Social Networking, Education, Marketing, Entertainment, Business, Shopping, and so many other things for making life easy. Now days, Social media craze is mounting to heights of success for every individuals. Youngsters are the most common users and most among of them are Students. Students just comment, share, like and post their feelings on social media like Twitter, Facebook and Youtube. Students feel free to discuss and share their experiences on social media in informal and casual manner without considering spellings and accurate grammar. Social media data are unstructured. These social media provides so many useful knowledge and information about the students' emotions, feelings, experiences and struggles in their studies outside the classroom. So following these Students track on the social media and it's an interesting outlook for educational researchers and practitioners to understand student learning experiences outside the classroom. This understanding will uncover so many unknown experiences which was not cleared or considered while the classroom discussion. This understanding about their experiences provides so useful data for the betterment of student in decision making, enhancing students' education quality, training and placement, withholding and achievements.

The amount of social media data provides chances to understand students' experiences, but their methodological difficulties to use social media data for educational purpose. In classroom studies, to know each and every students' point of view surveys, reviews, group discussions, interview and concealing was done. The innovative concept of using social media data focuses on the required information and knowledge to be extracted for educational purpose by understanding students' experiences.

The social media data like the students' comments, posts and feelings have to be used for understanding students' learning experience with the research goals:

- 1) To find and classify students problem in their learning.
- 2) Track students' good or bad experiences. Mining the social media data like students emotions will result to classify the group of students' according to their experiences and identify their problems to be solved to improve the education quality.

Several mining techniques for text document are used for mine various textual matches through the social area. Certain Social site applications with social websites like Facebook, Twitter, and Youtube gives so many chances to ascertain communication via people leading to communal learning and circulation of important knowledge such as comments, posts, emotions, along with likes.

Mining text oriented files is the acquaintance breakthrough technique that provides computational brainpower and this technique comprises of multidisciplinary fields, such as information rescue, text investigation, natural language dispensation, and in sequence classification based on logical and non-petty matches from colossal data sets. Users do not care about spelling mistakes and grammar rules. For text mining classification and clustering which has been extensively used for the examination of the shapeless text available on the large scale systems like Social area there are two approaches.

1.1 Problem Definition

Educational researches have been using old ways such as surveys, interviews, focus groups, classroom activities to collect data related to students' learning experiences. That will reducing scalability problem. As optimistic 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. There is no research found to directly mine and analyze student-posted content with considering the students' problem from uninhibited spaces on the social web with the clear goal of understanding students' learning experiences. The existing work has not measured student academic performance to identify the students' problem and classify them accurately for enhancing E-learning experiences.

1.2 Motivation

Many times, students' gets shy or afraid of clearing their problem during the classroom and this social media help them to just post whatever they feel at that time about their emotions. The schools and departments have been struggling with student recruitment and withholding issues. Graduates play an important role in nation's future workforce and which directly effects the nation's economic growth and global competency. The concept of combining Students' learning experiences for enhancing E-learning experiences is innovative for improving the training style or teaching style by which student to correct them at some required time without continuous concealing or surveying. 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[1]. Student can be trained or improve education quality as they have been classified. Students' learning experiences from social media will save the time to collect the data manually.

1.3 Objective

- To separate out meaningful information from the label "other".
- To classify student based on content shared in the social media.
- To understand issues and problems students encounter in their learning experiences.
- To classify new label "Good things" for students' combining with their emotional aspects and performance grade.
- Improved the "Probability of the labels and keywords" for new introduced labels.
- It can be used to inform educational administrators, practitioners and other relevant decision makers to gain further understanding of students' college experiences.

II. Literature Survey

2.1 Social Media

With the intensification of the Internet communiqué techniques, the World Wide Web have become a extremely imperative display place for users for a interaction with each other. In the course of this display place, users could share with a ease and broaden in sequence and dreams to any one and from anywhere.

Next to with the people's contacts with each other via these all social networking websites, progressively data and un-like in sequence information have been aggregated. Surrounded by all different information on these social networking websites, a quantity of has been considered as negative ones because they could be utilize to attack and to malign. Twitter is an exceptionally trendy micro blogging site, where users rummage around for appropriate and social information referred as breaking news, posts about celebrities, and trending topics. Users send short transcript messages referred as tweets, have limitation of 140 characters by length and could be view by user's followers. A follower is any person called when he is choosing to have other's tweets posted on one's timeline. For real-time, Twitters have been worn as a channel information giving out and it has been used in various brand elections, campaigns, and likely to a news media.

2.2 Naive Bayes Classification

The Naive Bayes Classifier technique has been based using Bayesian theorem and the best well-matched to the dimensionality of the inputs has sky-scraping. Over and over again outperforms more sophisticated classification methods, yet it is simple to operate. For the models, Maximum Likelihood estimates

all the parameters. To estimate the parameters, it has been require of small number of training. It operates glowing and powerfully in supervised learning.

In Bayesian analysis,

- Prior Probability: It is a belief and based on previous experience. It is a ratio of number of single objects and number of total objects.
- Likelihood: To classify a new object that this object belongs to which case.
- Posterior Probability: The final classification is made by combining both sources of information i.e. Prior and Likelihood to form a Posterior Probability using Bayes rule.
- Posterior Probability of X being a object α Prior Probability of total objects \times Likelihood of X given objects.

Above theorem guide us to solve the conditional probability of contrary and independent events. The estimation of the probability for an opinion could be containing positive, negative or neutral. It would produce good results .Conclusion is that Naive Bayes perform glowing in sky-scraping dependant features and outperforms regularly compared to Neural Networks, Decision Trees etc.,. Nevertheless criterion maximum likelihood parameter learning for Naive Bayes classifier tends to be suboptimal.

2.3Text Mining

The data mining done on social media data covers many uncover features of the social media or the social web i.e. Twitter, Facebook and Youtube. Mining file containing text has been known as Intelligent Text breakdown or acquaintance discovery in Text or it can said that Text Data Mining which can be used to mine the social media data. Mostly the social media data are unstructured format and to retrieve information from that is complex due the massive information. So, it requires specific processing methods and algorithms to extract useful information from that social web data.

2.4 Comparisons between Multiclass and Multilabel Classification

Table 2.1 Comparison between Multiclass and Multilabel classification

	MULTICLASS CLASSIFICATION	MULTILABEL CLASSIFICATION
PURPOSE	This means a classification task with more than two classes but not at the same time.	This means a classification task with more than two classes at same time.
PROPERTIES	It makes the assumption that each sample is assigned to one and only one label.	It is a predicting property of a data point.
NATURE	They are mutually exclusive.	They are not mutually exclusive.
EXAMPLE	A fruit can be either an apple or a mango, but not can be apple or mango at the same time.	A text can be at any document at same time.
NOTATION	Examples : $D = \{x_1, \dots, x_n\}$ Labels : $L = \{L_1, \dots, L_m\}$ Each example is associated with one label ($x, l \in L$)	Examples : $D = \{x_1, \dots, x_n\}$ Labels : $L = \{L_1, \dots, L_m\}$ Each example is associated with a subset of labels ($x, S \subseteq L$)

2.5 Study of Papers

SR NO	TITLE	AUTHOR	DESCRIPTION	PUBLI-CATION & YEAR
1	Understanding Customers Using Facebook Pages: Data Mining Users Feedback Using Text Analysis	Hsin-Ying Wu, Kuan-Liang Liu and Charles Trappey	For start-ups companies, Following steps are done in this paper: <ul style="list-style-type: none"> • Key phrases extraction and summarization. • Clustering algorithm: K-means. • Clustering quality evaluation: cohesion and separation. 	IEEE, 2014
2	Mining Social Media Data for Understanding Students' Learning Experiences	Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan	<ul style="list-style-type: none"> • Developed a workflow to integrate both qualitative analysis and large-scale data mining techniques. • Used a multi-label classification algorithm to classify tweets reflecting students' problems. 	IEEE, 2013

3	Content Matters: A study of hate groups detection based on social networks analysis and web mining	I-Hsien Ting, Shyue-Liang Wang, Hsing-Miao Chi and Jyun-Sing Wu	Two groups hate and like group. <ul style="list-style-type: none"> • Facebook Data Extraction • Data Pre-processing • Social Network Structure Features Extraction • Content Features Extraction by TF-IDF. • Group vectors 	IEEE, 2013
4	Recognizing User Identity in Twitter Social Networks via Text Mining	Sara Keretna, Ahmad Hossny and DougCreighton	They followed the steps: <ul style="list-style-type: none"> • Input: unlabeled tweets. • Feature extractor: feature vector is extracted • Classifier: to predict identity of user. 	IEEE, 2013
5	Twitter Trending Topic Classification	Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md. Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary	<ul style="list-style-type: none"> • They classify Twitter Trending Topics into 18 general categories such as sports, politics, technology ,etc. • They experiment with 2 approaches for topic classification; <ul style="list-style-type: none"> (i) the well-known Bag-of-Words approach for text classification (ii) network-based classification. 	IEEE, 2011

III. Proposed Work

3.1 Overview of Proposed Work

Proposed work focus on the students' emotions express on the social site by which they can be classify into good and bad experiences. The proposed work is only the first step towards enlightening actionable insights from student-generated content on social media in order to pick up education quality. The proposed system aims at which are related to engineering students' Twitter posts that are to understood arguments and problems for their educational experiences. Initially to be conducted was qualitative investigation on such a gathered sample taken from tweets which are related to engineering students' belongs to college life. The proposed system finds engineering students bump into problems are as follows: lack of social engagement, heavystudy load, and sleep dispossession. Considering the outcomes to be brought the proposed system will implement of multi labelled classifiers algorithm to identify and organize tweets which will be reflecting students' problems. The proposed system has been used for the supervised learning technique for the organization of classified the twitter tweets and such a Foremost it has to be employed a healthy known text arrangement technique called Naive Bayes (N.B.) [10]. A document for NB would be modelled for the presence and absence of particular words. A variation of NB is Naive Bayes multinomial (NBM) and this will considers the rate of recurrence of words and could be expressed as follows:

$$P(c|d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t_k|c)$$

In which it represents as : P (c | d): the probability of a document as d individual in class as c, P (c): the prior probability of a document taking place in class as c, and P (t k | c): the provisional probability of term as tk happening in a document of class as c. A document as d in such a case has been trending defined or tweets associated with tags such as: #hashtag.

The proposed system has been sequences of steps which will recognize the social media data for educational purposes which would overcomes all major limitations of both manual qualitative investigation and large scale computational investigation of user-generated textual at ease There can be no pre-defined categories of the data, subsequently the proposed system needed to explore what all the students are saying in the tweets and what they want to say.

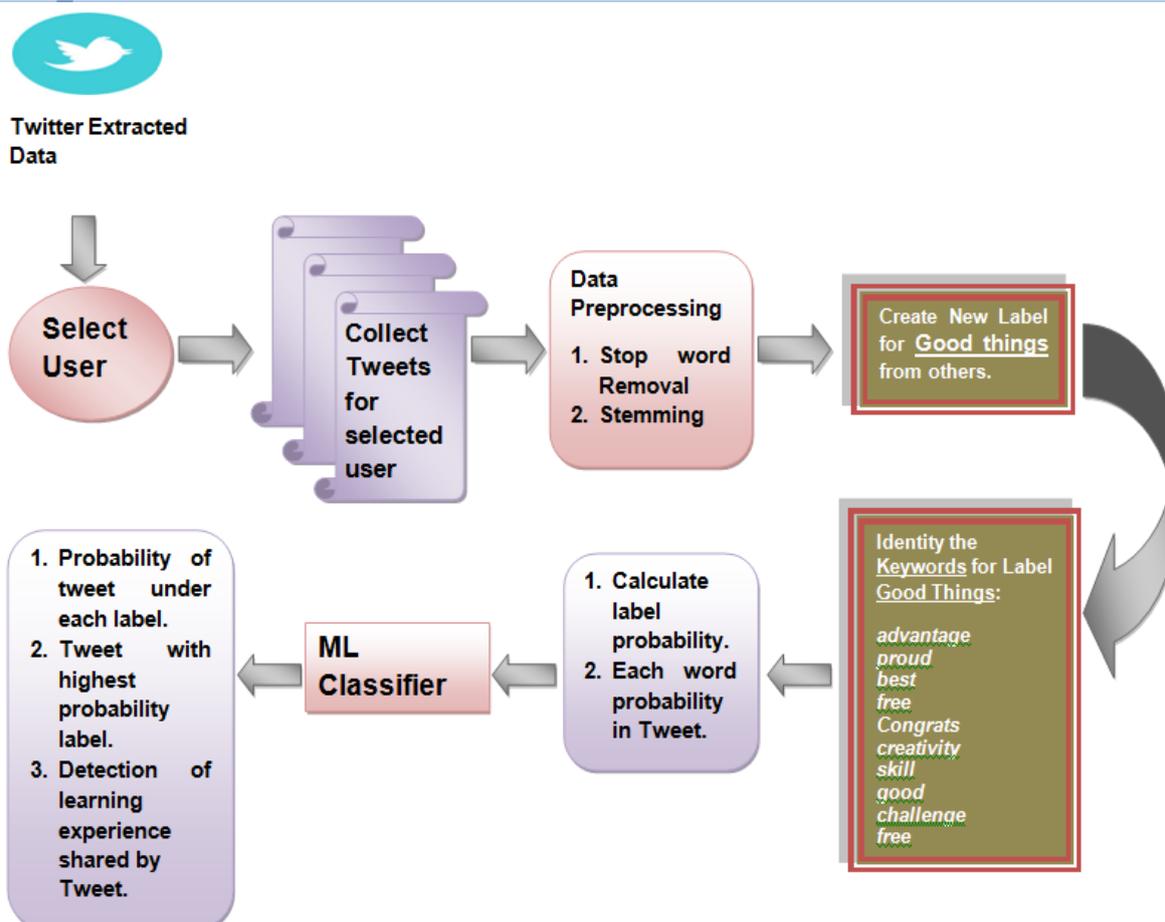


Fig.3.1 Architecture Diagram

As a consequence proposed system first conducted an inductive content analysis on the engineering Problems dataset. Inductive content analysis is one popular qualitative research method for manually analyzing text content. The proposed system extends 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. Existing system have five prominent themes as under by feature extraction:

Table 3.1 list of category with their keywords[1]

CATEGORY	MOST PROBABLE KEYWORDS THAT OCCURS IN TWEETS
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, disgusting
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, tire, 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, nosleep
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

In Proposed system, we used six existing prominent themes or labels with new theme or label: "GOODTHING". The most probable features extracted under new label from analysing the tweets are as follows:

Table 3.2 Details of new Label with keywords

CATEGORY	MOST PROBABLE KEYWORDS THAT OCCURS IN TWEETS
Others	All rest words are considered to be in this category
NOTE: Now separating out some useful words which will be coming under new label "GOODTHING" by analysing the extracted tweets and then separated out the keywords for this new label.	
GoodThing	Advantage, proud, best, free, Congrats, creativity, skill, good, challenge, free, happy, early, like, love, relax, easy, fun, excite, progress, thank, god, lucky

3.2 Flow Chart of Proposed Flow.

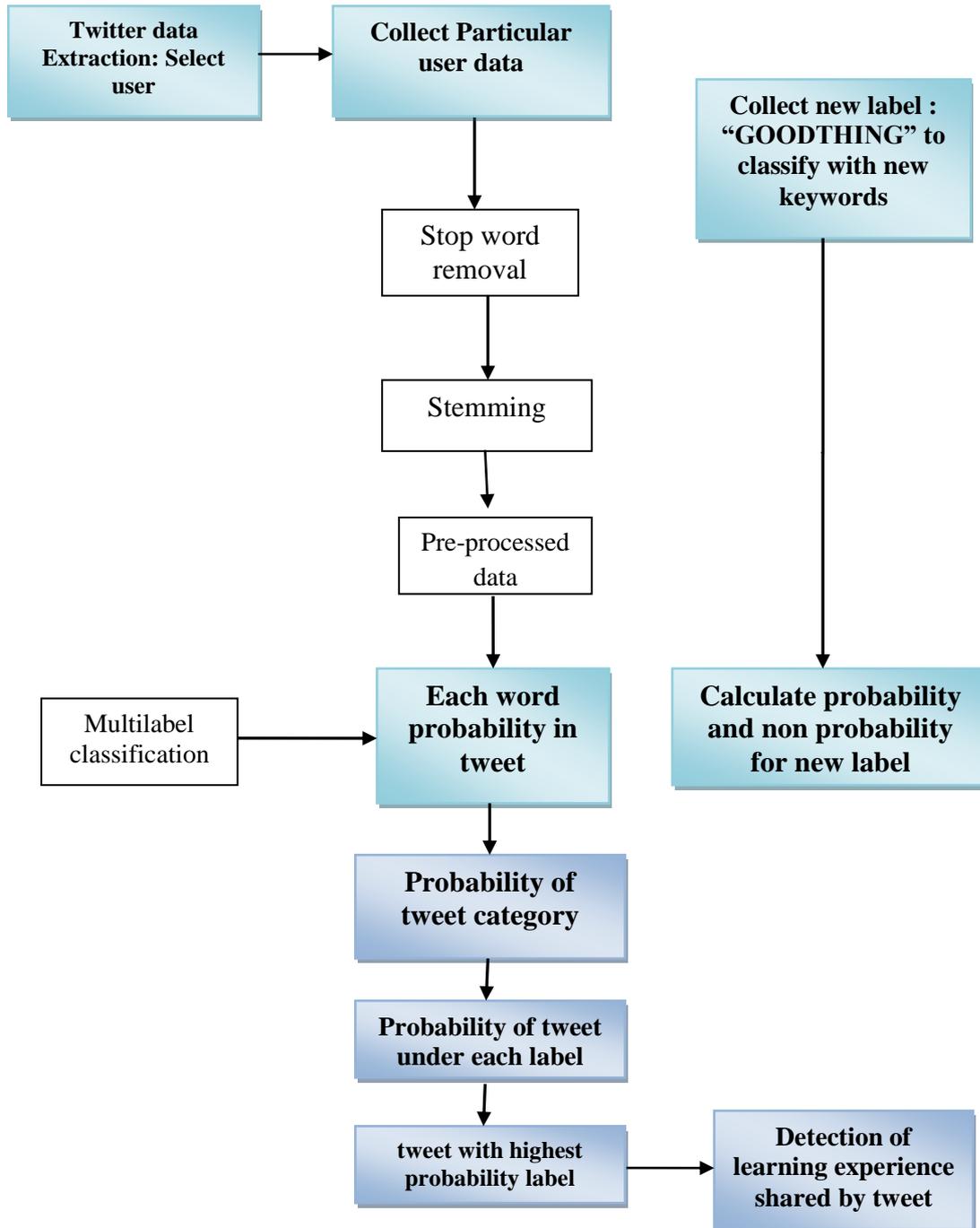


Figure 3.2 Flow chart of proposed work

IV. Experimental Results

This section describes several experimental evolutions on inputted Student tweets on twitter extracted from import.io specific dataset.
Existing system result

1st train data: No. Of tweets: 90, No. Of label: 6 and keywords: 215

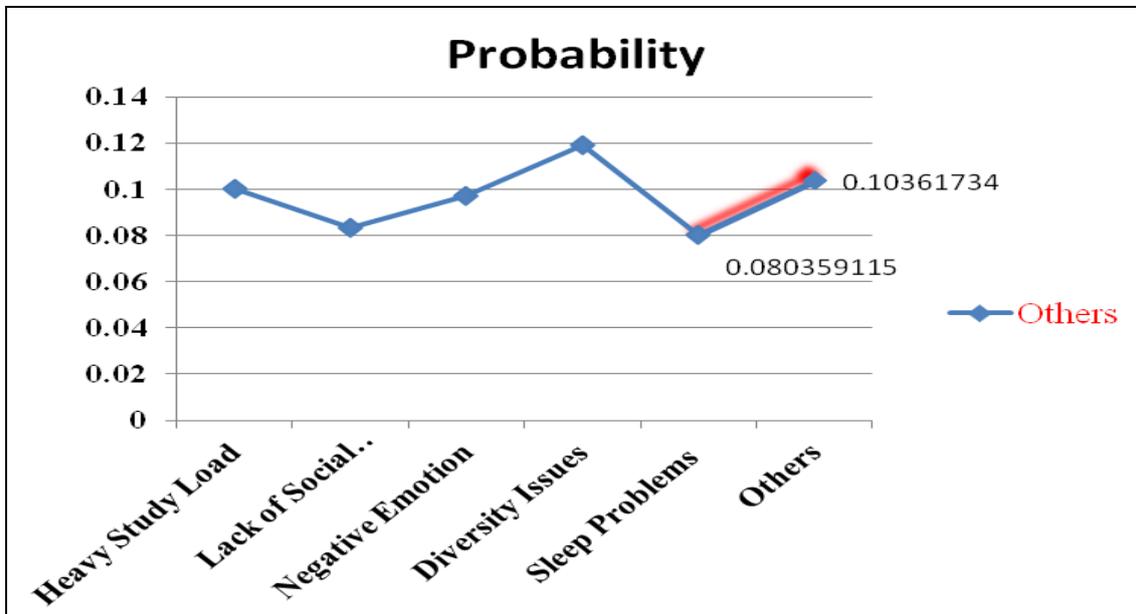


Fig 4.1 Probability of keywords falls in which label existing system

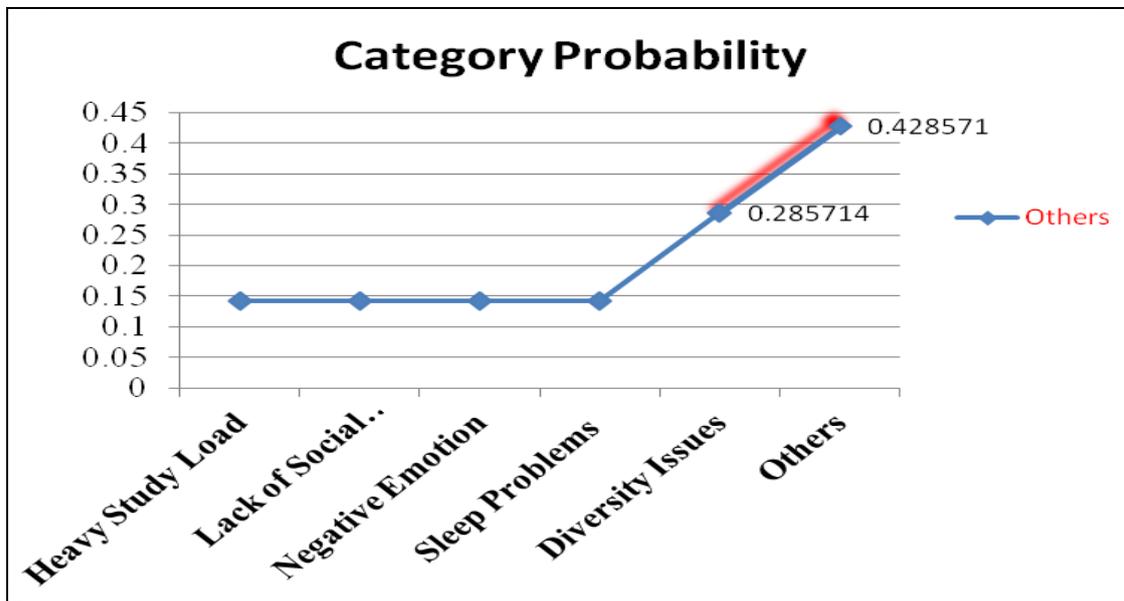


Fig 4.2 Result Category Probability of Keywords Labels existing system

Proposed system result

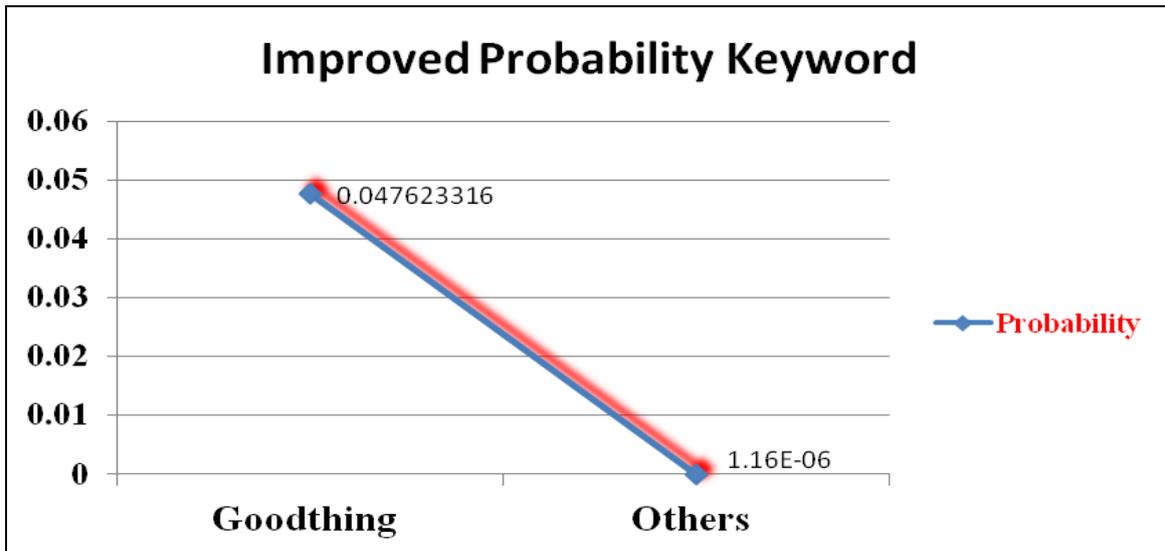


Fig 4.3 Improved Probability of "Goodthing" from "Others"

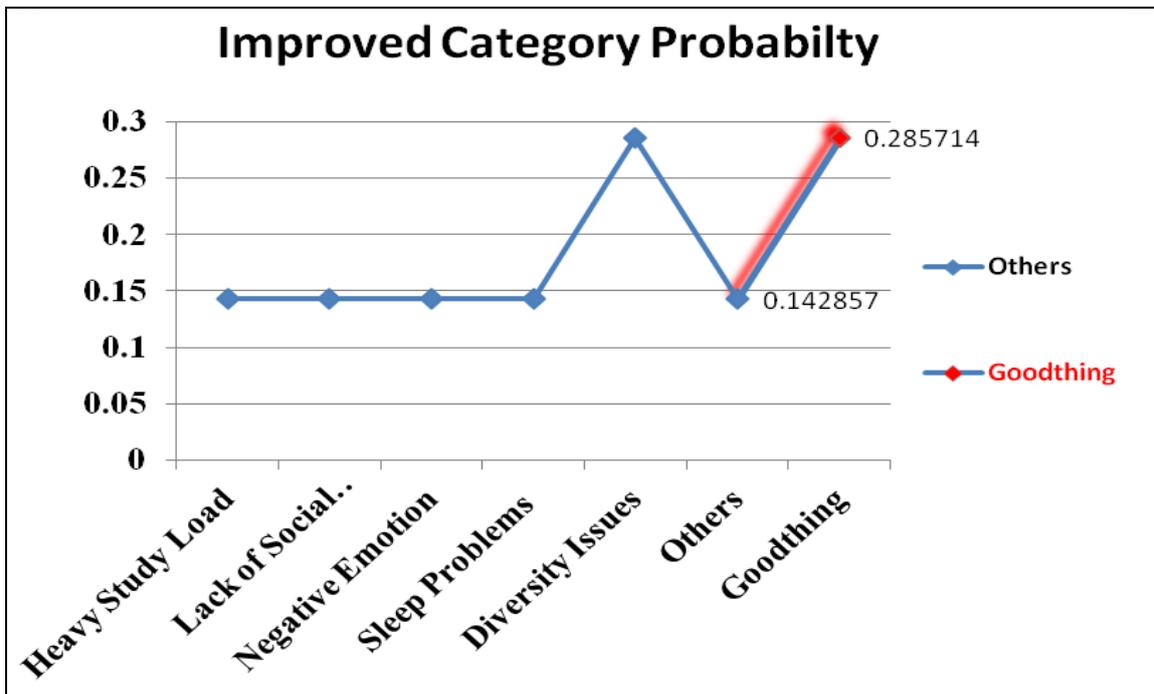


Fig 4.4 1st train data: Improved Category Probability of Keywords Labels

After Successfully Evaluate the result of Proposed Approach with compare to Existing Approach, we will see a comparative study of both system.

Table 5.6 Comparison Table of Proposed and Existing system

	Proposed System considering new label "Goodthing" and "Others"	Existing System
Dataset : (implemented)	<ul style="list-style-type: none"> • Real time Twitter data • 1st train data: 90 tweets, 215 keywords-Real dataset • 2nd train data:200 of tweets, 215 keywords-Real dataset 	Only Prudue University.
Parameter:	<ul style="list-style-type: none"> • Probability of keywords. • Probability of labels. (Category and non-category) 	Probability of tweets under five labels.
Labels :	Seven Labels: HeavystudyLoad Lack of Social-Engagement Negative Emotion SleepProblems Diversity Issues Others Goodthing	Only consideration of five labels and "Others" label as extra case to fetch unwanted tweets.
Extractor:	Use import.io extractor. (collect large tweets monthly)	Prudue University

V. Conclusion

The conclusion to the research done till now is those researchers in learning analytics, educational data mining, and learning technologies can be very advantageous. 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.

From study of various labels for classifying student learning experiences' many possibility of falling students under that different category classified as per the analysed result, it is cleared that after implementing proposed work many of the useful tweet which were going under the others and not considered. That all real tweets were collected, analysed and found the probability of that tweets falling under "Goodthing". By using Naive Bayes probability rules, we classified the tweets for finding the Probability of the keywords and probability tweets under labels.

From studies of different papers, only knowing the sentiment of student shared tweets does not provide much actionable knowledge on relevant interventions and services for students. In sentiment analysis, tweets can be classified positive, negative or neutral but cannot be judged that it belongs to which category. So, in the proposed system multi label classification model was used which will make fall tweet into multi-categories at same time.

The proposed system successfully improved the probability and category probability of the two labels: "Others" and "Goodthing" labels. And concluded that student just don't only post their bad experiences' but also good experiences' on social media. And even used a new extractor import.io which helps in making database and dataset and can be used as web crawler.

References

- [1]. Xin Chen, Mihaela Vorvoreanu, and Krishna Madhavan "Mining Social Media Data for Understanding Students' Learning Experiences," IEEE transactions on learning technologies, manuscript id 1, pp. 1-14, 2013.
- [2]. Ajay Kumar Pal, Saurabh Pal, "Classification model of Predication for Placement of Students," in I.J. modern Education and Computer Science, pp. 49-56, November 2013.
- [3]. Hsin-Ying Wu, Kuan-Liang Liu and Charles Trappey, "Understanding Customers Using Facebook Pages: Data Mining Users Feedback Using Text Analysis", IEEE 18th International Conference on Computer Supported Cooperative Work in Design, 2014, pp. 346-350.
- [4]. I-Hsien Ting, Shyue-Liang Wang, Hsing-Miao Chi and Jyun-Sing Wu, "Content Matters: A study of hate groups detection based on social networks analysis and web mining ", IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013,pp. 1196-1201.
- [5]. Sara Keretna, Ahmad Hossny and DougCreighton," Recognising User Identity in Twitter Social Networks via Text Mining", IEEE International Conference on Systems, Man, and Cybernetics, 2013 , pp. 3079-3082.
- [6]. PAPER 5: Kathy Lee, Diana Palsetia, Ramanathan Narayanan, Md. Mostofa Ali Patwary, Ankit Agrawal, and Alok Choudhary, "Twitter Trending Topic Classification", 11th IEEE International Conference on Data Mining Workshops, 2011 , pp. 251-258.
- [7]. N. Anitha, B. Anitha, S. Pradeepa, "Sentiment Classification Approaches – A Review," in International Journal of Innovations in Engineering and Technology, vol. 3, Issue 1, pp. 22-31, October 2013.
- [8]. Nitin Pawar, Prof. S. K. Sonkar, "An Approach Towards E-Learning Using SVM Classification Technique and Ranking Technique in Microblog Supported Classroom: A Survey," in International Journal of Emerging Technology and Advance Engineering, Volume 3, Issue 7, July 2013.
- [9]. A. K. Santra, S. Jayasudha, "Classification of Web Log Data to Identify Interested Users Using Naïve Bayesian Classification," in International Journal of Computer Science Issues, Vol. 9, Issue 1, pp. 381-387, January 2012 .
- [10]. C. D. Manning, P. Raghavan, and H. Schtze, "Introduction toInformation Retrieval." New York, NY, USA: Cambridge University Press, 2008.