Mining Social Networks for Analyzing Students Learning Experience and their Problems

Blessy Geo.V.M[1],

[1]Research Schoar,Department of Computer Science,
Vels University, Pallavaram,
Chennai-600043.

Dr. S.Prasanna[2]

[2]Associate Professor, Department of MCA,
Vels University, Pallavaram,
Chennai-600043.

Abstract - **On different social media sites, students discuss and contribute to their daily encounters in a casual and informal manner. Analyzing such data, though, can be difficult. The complication of students' experiences reflected from social media content requires human understanding problem of a student's experience exposes from social media sitedrequirehuman investigation or communication.Examining data from such a social media can be demanding task. In this article, we develop a workflow to combine both qualitative analysis and large-scale data mining techniques. It pays a concentration on engineering student's Twitter posts to recognize the problem and the difficulty in their educational practices. Based on this result, a multi-label classification algorithm that is Naive Bayes Multi-label Classifier algorithm and Memetic classifier is applied to categorize tweets presenting students' problems.Memetic classifier is a population based approach to split individuals education for problem search, which has had their own advantages in solving optimization problems.**

Keywords: Naïve Bayes Multi-label Classifier, Memetic Classifier, Social media, Statistical Analysis.

I. INTRODUCTION

A data mining study has effectively created several methods, tools, and algorithms for supervision huge amount of data in response to real-world difficulty. As social media are broadly used for a variety of purposes, enormous amounts of user shaped data are near and can be made accessible for data mining. Data mining of social media is able to enlarge ability of accepting innovative experience, the apply of social standard and increase business intelligence to here good services and enlarge innovative opportunities. The main purpose of the data mining method is to collectively hold large-scale data, take out actionable patterns, and get insightful knowledge. Social media sites such as Twitter, a Facebook present grand place for students to share pleasure and struggle, sentiment and strain, and gain communal support.The rising field of learning analytics and learning, data mining is listening carefully on analyzing structured data obtained beginning Course Management Systems (CMS), classroom knowledge usage, or prohibited online learning environments to notify educational resolution making. However, to the greatest of our knowledge, there is no research, establish to directly mine and examine student- posted content from unrestrained spaces on the social web with the obvious goal of understanding students' educational experiences.

The research aim of this study is 1) to reveal a workflow of social media data sense-making for learning purposes, integrating equally qualitative analysis and large-scale data mining techniques to discover engineering students' casual conversations on Twitter, in order to understand the issues and troubles students meet in their learning experience.

II. LITERATURE REVIEW

Web (data) mining is one of the intelligent computing techniques in the context of Web data management. In general, Web mining is the means of utilizing data mining methods to induce and extract useful information from Web data information. Web mining research has attracted a variety of academics and engineers from database management, information retrieval, artificial intelligence research areas, especially from data mining, knowledge discovery, and machine learning,etc. Basically, Web mining could be classified into three categories based on the mining goals, which determine the part of the Web to be mined: Web content mining, Web structure mining, and Web usage mining .

Learning analytics and educational data mining (EDM) 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. Educational Data Mining (EDM) is the

application of Data Mining (DM) techniques to, its objective is to examine this type of information in order to determine educational research problems.

III. RELATED WORKS

The theoretical basis for the importance of informal data on the network can be strained from Goffman's theory of social presentation. Goffman's theory of social presentation is widely used to give particulars on the web nowadays and it is generally used to explain mediated communications on the web today. Student's online conversation reveal aspects of their experience that are not easily seen in proper classroom settings, thus are regularly not documented in learning literature. Below are reviews of studies on Twitter from the fields of data mining,natural language processing and machine learning. These studies have extra importance on statistical models and algorithms. They cover up a wide range of topics popularity calculation, topic discovery, event detection and tweet classification. Amongst these topics, tweet classification is mainly relevant to this learning. Popular classification algorithms contain Naïve Bayes, Logistic Regression, Decision Tree, Maximum Entropy and Support Vector Machines (SVM).

IV. PROBLEM DEFINITION

Generally, educational researchers contain utilizing routines such as survey, classroom exercises, interviews, and to collect information identified through students learning experience. At the end when energized on their experience, students require to think what they were viewing and doing in the history which might contain gotten to be ignored. Twitter tweets of Students contain enormous of data which can give a variety of activities with them. By receiving the data we can predicate the troubles of the students. In the existing system Naïve Bayes Multi-label Classifier is used, this classifier gives a good outcome but it's extremely time consuming. So, to defeat this limitation,we proposed a method called as a "Memetic Classifier" based on genetic algorithm. By using Memetic Classifier, we are capable to recognize the students' problems accurately and classified.

V. PROPOSED METHODOLOGY

The research aim of this study is to reveal a workflow of social media data mining for learning purposes and to discover engineering student's informative conversations on Twitter, through which we appreciate the issues and troubles students come upon in their knowledge experiences. First a sample is taken from student and then it conducts qualitative investigation of that sample which is related to engineering students educational life. It found engineering students come upon problems such as heavy learning load, lack of social meeting, and sleep absence. Stand on these outcomes, the authors apply a multi-label classification algorithm to classify tweets presenting students' problems. Then MemeticAlgorithm ispractical to make, more perfectresult, itwill filter after that used the algorithm to organize a detector of student problems.

The significantposition in proposed study is, First, it intends a workflow to association and incorporate a qualitative investigatethe methodology and large scale data mining techniques. It bases our data-mining algorithm on qualitative approachingconsequential from human perception, so that it can gain deeper accepting of the data. Then apply the algorithm to another large-scale and unfamiliar dataset, so that the physical method is improved. Second, the paper provides bottomless insights into engineering students educational experiences as reacting in casual, uncontrolled environments. Many issues and troubles such as study-life balance, lack of sleep, lack of social engagement, and lack of variety clearly materialize. These could bring attentiveness to educational examiner, policy-maker.

A.Naive Bayes Multi-label Classifier

We built a multi-label classifier to classify tweets based on the categories developed in earlier for the ease examination stage. For the task, multi label classification, Nave Bayes Classifier be invented to be modified to multi label data. For multilabel classification, the situation is slightly more complicated, than the other classifier because each text gets assigned multiple labels. Data generated by students in Social media.
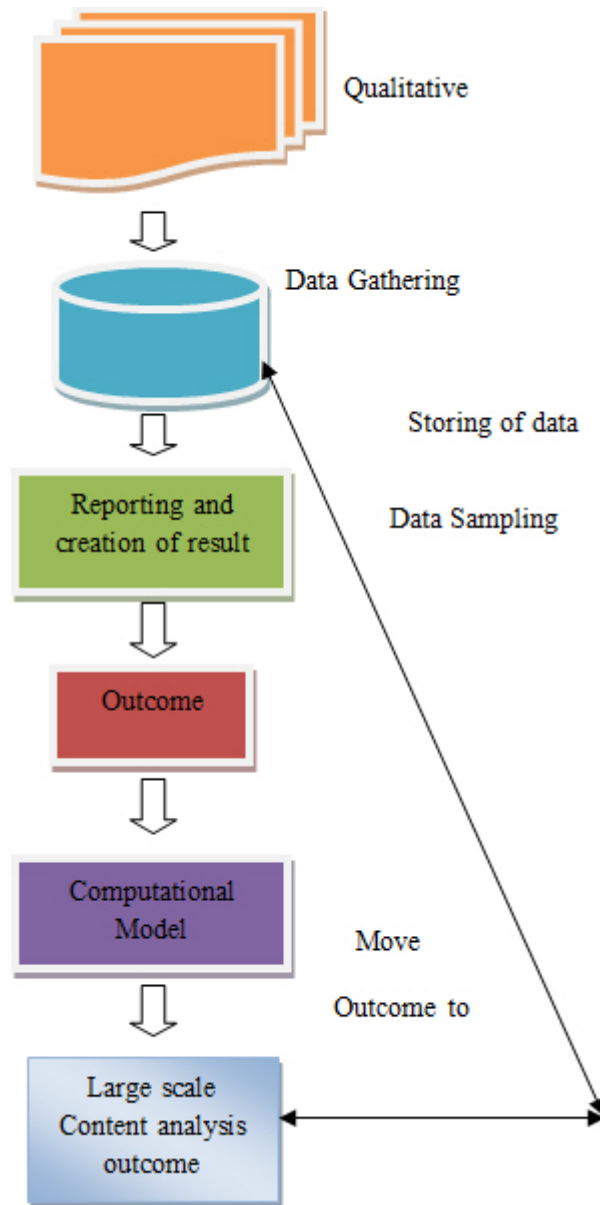
Fig.1. System Workflow

Naive Bayes executes well in many complicated real-world difficulty. Even though it is often outperformed by other techniques such as boosting trees, Max Entropy, Support Vector Machine,etc, Naive Bayes classifier is tremendously efficient since it is less computationally and it requires a small amount of training information. One well-liked way to perform multi-label classifier is to convert the multi-label organization problem into multiple single-label classification problems.

Suppose there are a total number of N words in the training document collection (in our case, each tweet is a document) W ¼ fw1; w2; ::::;wNg, and a total number of L categories C ¼ fc1; c2; ::::; cLg. If a word wn appears in a category c for mwnctimes, and appear in categories other than c for mwnc times, then based on the maximum likelihood estimation, the probability of this word in a specific category c is

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

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

$$P(w_n|\ c') = \frac{m_{wnc'}}{\sum_{n=1}^{N} m_{wnc'}} (2)$$

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

$$P(c) = \frac{C}{M} \qquad (3)$$

and the probability of other categories c' is

$$P(c') = \frac{M-C}{M} (4)$$

B.MemeticClassifier

The Memeticclassifier is a population based move toward and it's agrowth of genetic algorithm. There are mostly used for optimization. The MemeticClassifier is comprehensive by a local explore. The genetic algorithm begins with a randomly selected population of chromosomes and the population is initialized at casual or using a heuristic. The Memeticclassifier refers to hybrid algorithms, a wedding between a population based global technique and a local explore made by every of the individuals. The innovative implemented or developed Memetic classifier is an addition of a Memeticclassifierby a classification technique. After the accomplishment of the optimization process of Memetic algorithm applying or appending a classification method to it. This Memetic classifier proceeds a completely optimized classification consequence in very less time as evaluate to Naïve Bayes multi-label classifier. Quite often, MA is also referred to in the text as Baldwinian evolutionary algorithms (EA), Lamarckian EAs, cultural algorithms, or genetic local search.

VI. CONCLUSION

Mining social media data is cooperative to researchers in learning analytics, educational data elimination, and learning skill. It gives a way to investigate social intermediate statistics that defeat the main limitations of both corporal qualitative analysis and huge scale computational study of user formed textual contented. From this work done and survey of Interpreting Students Behavior is cooperative to find the drawbacks of the accessible classification algorithm. The newly implemented Memetic Classifier hasanalyzed the very large amount of data in a short period of time. This newly developed system provides the classification or clustering of student problems and issues and also provides the opinion analysis of the database. As per the survey and work done this system very useful in institutes, organizations, universities. We can enhance in image processing like images, emoticons, videos,etc. This system is also very useful for industry, manufacturing companies, banking sectors, government sectors etc. in future for identification of employees' actions, their behaviors, product feedback, for banking feedback and related to their services.

REFERENCES

[1]   R. Ferguson, "The state of learning analytics in 2012: A review and future challenges," KnowledgeMediaInstitute, Technical Report KMI-2012-01, 2012.
[2]   J.M. DiMicco and D.R. Millen, "Identity Management: Multiple Presentations of Self inFacebook," Proc. The Int'l ACM Conf. SupportingGroup Work, pp. 383-386, 2007.[3] M. Vorvoreanu and Q. Clark, "Managing Identity Across Social Networks," Proc. Poster Session at the ACM Conf. Computer SupportedCooperative Work, 2010.
[3]   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. 252-
[4]   X. Chen, M. Vorvoreanu, and K. Madhavan, "A Web-Based Tool for Collaborative Social Media Data Analysis," Proc. Third Int'lConf. Social Computing and Its Applications, 2013.
[5]   G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining Multi-label data,"Data mining and knowledge discovery handbook, pp. 667-685, 2010.
[6]   "Using the Twitter Search API | Twitter Developers," https://dev.twitter.com/docs/using-search, 2013.
[7]   J. Srivastava, R. Cooley, M. Deshpande, and P.-N.Tan.Web usage mining: discovery and applications of usage patterns from web data.SIGKDDExplor.Newsl., 1(2):12–23, 2000.
[8]   Pallavi K. Pagare.(2014) Analyzing Social Media Data for Understanding Student's Problem. International Journal of Computer Applications (0975 – 8887) Innovations and Trends in Computerand Communication Engineering (ITCCE-2014.
[9]   Heyam A. Al-Tarawneh.(2014) The Influence of Social Networks on Students' Performance Vol. 5, No. 3 March 2014 ISSN 2079-8407 Journal of Emerging Trends in Computing and Information Sciences
[10] L.R.JeevithaR.Priyanka.(2015) 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.
[11] Gauri Joshi1, Mr. SamadhanSonawane.(2015) Filtering and Classification Of User Based On Social Media Data Using Memetic and Naïve Bayes Methods International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4795-4798.
[12] B.Srinivasa (2014) Mining Social Networking Data for Classification Using ReptreeVolume 2, Issue 10, October 2014 International Journal of Advance Research in Computer Science and Management Studies.