ANALYZING STUDENT'S LEARNING EXPERIENCES THROUGH SOCIAL MEDIA DATA USING MACHINE LEARNING TOOL

[1]AJINKYA .A. LONDHE, [2]JAMGEKAR R.S, [3]SOLUNKE B. R

[1]M.E (Computer) [2]M.E (Computer), Assistant Professor, [3]M.TECH (Computer), Assistant Professor, Department of Computer Science & Engineering, N B Navale Sinhgad College of Engineering, Solapur 413255.

Abstract— Social media which can be stated as online media supporting social interaction & user contribution is playing a crucial role in social networking and sharing of data. School, Colleges and universities are beginning to accept social media as a source of mean for enhancement of education system. Student's comfortable and accidental talk's on social media shade light into their educational experience, mind-set, and worries about their learning procedure. Social media sites such as twitter, Facebook, and you-tube provides grand platform to large amount of user without any restrictions to share their opinions, educational learning experience and concerns via their posts Assessment of such data in social network is quite a challenging process. In the proposed system, there will be advancement to mine the data which constitute both qualitative analysis and extensive data mining technique. Tweets will be categorized into different categories considering various striking themes. We use WEKA a data mining/machine learning tool to integrate Naive Bayes classifier and support vector machine (SVM)on mined data for qualitative analysis purpose to get the deeper understanding of the data and obtain more accurate results out of the data-set using label based measure to analyze the results. This scheme, presents an approach that show how unconstrained social media data can provide awareness and insights into students' learning experiences.

Keywords— Computers and education, Social networking, Social media data, WEKA, Naive Bayes, SVM and Data mining.

I. INTRODUCTION

Research from the fields of data mining has admirably produced various technique, tools, and algorithms for managing huge amounts of data to answer real-world troubles. Social media is widely used for various purposes, vast amounts of user created data can be made available for data mining. Social networking is for everyone and it's now such an important and massive part of all our lives. According to the survey sharing of the data is high in the social sites like twitter and face book [12].It provides a great platform for students to express theirviews, stress, emotions, opinions, issues, joy, struggle, feelings and seek social support. Student's discuss and share their everyday encounters in formal and informal manner on different social media sites. This students' tweets and post can provide valuable and implicit knowledge that might be very helpful for an institution to understand the difficulties of the student he/she facing in the learning system. Thus, improving education quality, and thus enhance student recruitment, retention, and success [4].However this data or information is used to upgrade the education system of the particular institution. These social network data mining provide an scope to modify education system as students play a crucial role eventually to make an impact on nations economic growth. The abundance of social media data provides opportunities to understand students' problems, but also raises methodological difficulties in analyzing data for educational purposes. Hand operated analysis cannot deal with such huge amount of data, while automatic algorithms usually cannot capture in-depth meaning within the data [2]. There are so many traditional methods available such as questionnaires, surveys and face to face interviews to analyze the student's learning barriers in an educational institution [5]. But the main problem with these methods is these techniques are time consuming and cannot be performed effectively on regular basis as the analysis has to be performed manually. One more important problem is the data collected in such manner may not be reliable as the student may not disclose what they actually feel when compared to an informal medium like social media. Automated prediction of trends and behaviors: Mining automates is the process of finding predictive information in a large data-set

Sentiment analysis which is one of the existing system used for opinion mining which classifies the results as positive, negative and neutral. For understanding the tweet in a deeper manner this sentiment analysis is not sufficient. Therefore, our study requires a qualitative analysis, and is impossible to do in a fully unsupervised way. Sentiment analysis is, therefore, not applicable to our study. In proposed system, we used a multi-label classification model where we allowed one tweet to fall into multiple categories at the same time. The drawbacks are in our study, through a qualitative content analysis, we found that students are largely facing heavy study load, and are not able to manage it efficiently. Heavy study load results into other issues like sleep problems, and other psychological issues. Our work is only the first step towards revealing actionable vision from student-generated content on social media in order to improve education quality.

The research goals of this study are:

1) To set fort hand design a workflow of social media data analysis for understanding student learning experience problems and to take proper decisions to improve the education system of the institution.

2) To explore students' informal conversations on social media sites, in order to understand issues and problems students encounter in their learning experiences.

II. PROPOSED METHODOLOGY

We propose automation system in extracting and mining data, through the informal posts and chats on social media platforms, made by the students, in order to exactly know about their concerns and issues, on a larger scale. In this system, the students' data will be mined against certain standard data sets and several algorithms will be used in order to understand the relevance of their concern and feelings, through their posts or chats on the social media engine. In this project we intend to develop data mining system using Stop-word and stemming for pre-processing, Naive Bayes and SVM (Support Vector Machine) for classification to demonstrate the workflow of social media data sense-making for educational objective, fussing both qualitative analysis and various data mining techniques.

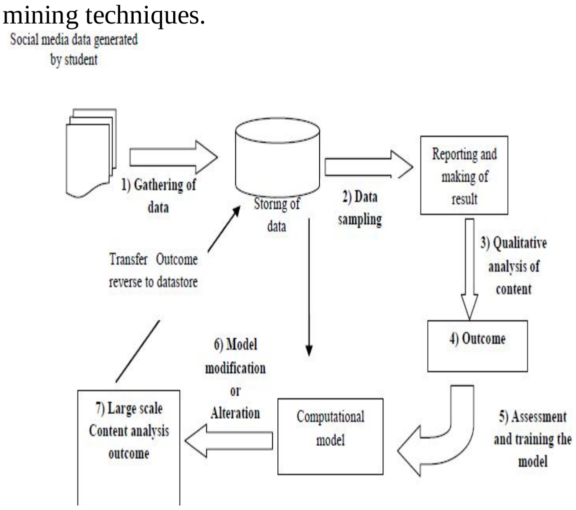

Fig 1.System architecture

System Flow:-

The system flow is shown in the figure above: In this system there is an investigative procedure to find the appropriate data.

- Gathering of tweets/post/data from various social media sites. This corresponds to the step 1 In Fig. 1.
- Next in the step 2 and 3 of Fig. 1. The inductive content analysis is perform on the sample of students tweets.
- In step 4, it is found that the major problem that comes into students fall into numerous well-known categories. Based on these categories, we used WEKA TO integrate multilabel Naive Bayes classification algorithm and SVM classifier is executed for classification.
- In step 5 the performance of the classifiers is estimated by comparing it with other multilabel.
- In step 6 the classification algorithm is applied by System to prepare a detector that help recognition of student's problems.
- The results are provided by step 7 help educators to identify at issues students are facing and make decisions on proper interference to preserve them and provide better education system.

III. EXPERIMENT AND RESULTS

STEP 1. The system provides a GUI for user to load the data-set on which data processing task needs to be performed to develop various prominent themes/categories.

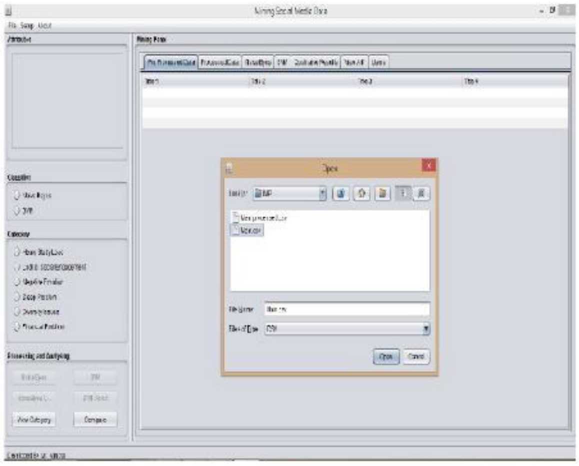

Fig 2. Loading sample dataset in CSV format.

STEP 2. After loading the dataset and data processing ,the developed categories ARFF files s need to loaded as WEKA need ARFF file format for computing the various dataset using Naive Bayes and SVM for analysis purpose.

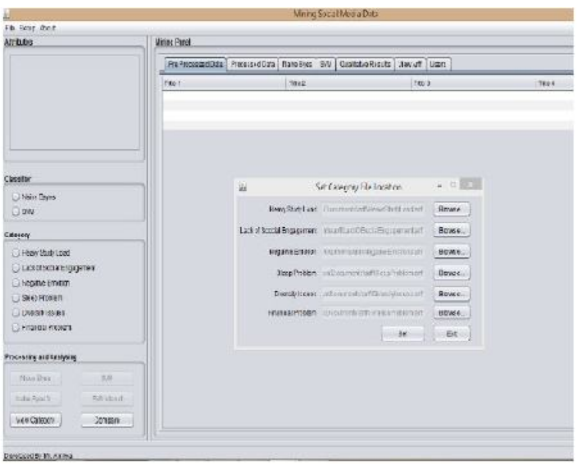

Fig 3. Loading of category ARFF file for analysis.

STEP 3. Select the algorithm using which you want to analyze the data-set. After selecting the classifier select the category and select view category to know the keywordsof that category. Select the classifier to compute the results.

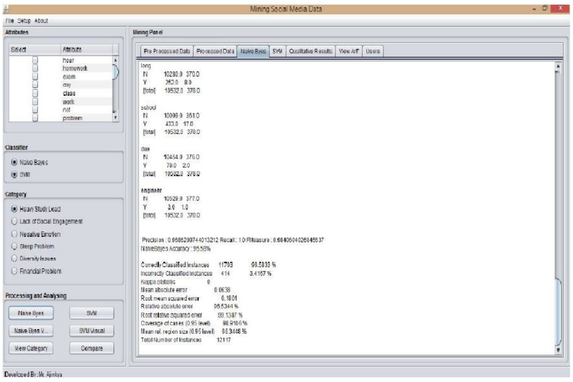

Fig 4. Evaluation of results.

IV. RESULTS AND COMPARISONS

1. Evaluation Measures for classifiers:

Typically used measures to evaluate the performance of classification models include accuracy, precision, recall, and the average between precision and recall—the F1 measure score. For multi-label classification, the situation is slightly more sophisticated, because here multiple labels are assigned to each document. Among these labels, some may be correct while others may be incorrect. Therefore, there are usually two types of evaluation measures—example-based measures and label-based measures [6].

1.1 Example-Based Evaluation Measures:

For a certain document D, suppose the set of true labels it falls under is P, and set of predicted labels by the classifier is Q, then for this document, accuracy is the correctly predicted number of labels divided by the number of labels in the union of P and Q. Precision is the correctly predicted number of labels divided by the total number of labels in Q, while recall is the correctly predicted number of labels divided by the number of true labels. Suppose there are a total of M documents {D1; D2….; DM} then the accuracy precision, recall, and F-Measure averaged over the M documents are

$$\text{Accuracy (a)} = 1/M\sum_{i=1}^{M}(Pi\cap Qi/Pi\cup Qi)\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

$$\text{Precision (p)} = 1/M\sum_{i=1}^{M}(Pi\cap Qi/Qi)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$\text{Recall (r)} - 1/M\sum_{i=1}^{M}(Pi\cap Qi/Pi)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

$$\text{F-Measure} = 1\backslash M\sum_{i-1}^{M}\frac{2.\,Precision * \frac{Recall}{Precision} +}{Recall}\ldots\ldots(4)$$

1.2 Label-Based Evaluation Measures:

From the categorized tweets, the system will display the tweets and the categorized section.Where TP is true positive, TN is true negative FN is false negative and FP is false positive[11]. For single category C,

Accuracy (a) = TP+TN / TP+TN+FP+FN…….. (5)

Precision (p) = TP / TP+FP…………………… (6)

Recall (r) = TP / TP+FN……………………….. (7)

F-Measure =2TP / 2TP+FP+FN………………. (8)

Naive Bayes classifier is extremely efficient since it requires less computation and a small amount of preparation information. Support Vector Machine (SVM) [9] is one of the most used and accurate classifiers in many machine learning tasks, But our Comparison experiment shows that SVM exceeds Naive Bayes in this study.

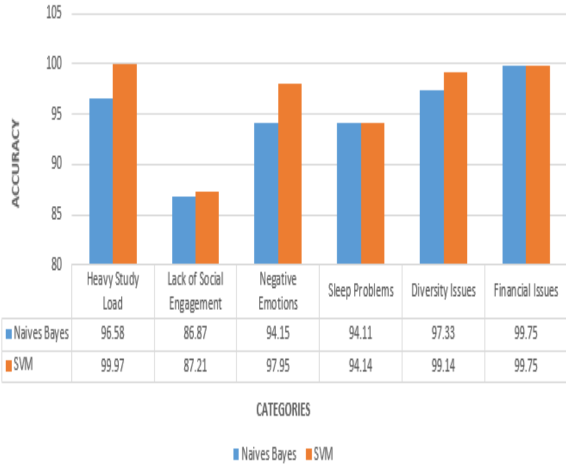

Fig 5.Label-Based Accuracy Comparison of Naive Bayes & SVM

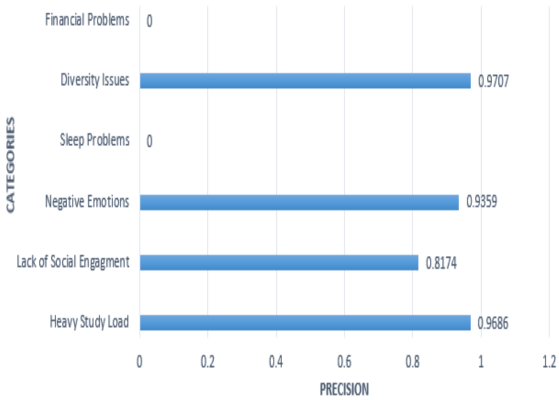

Fig 6.Label-Based Precision on Each Category

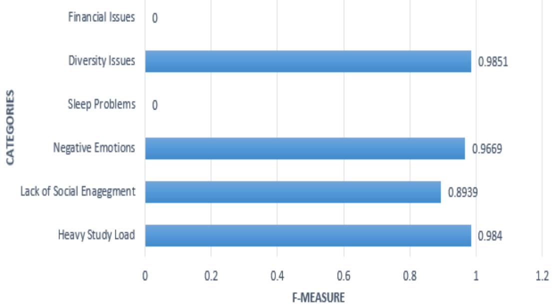

Fig 7. Label-Based F-Measure on Each Category

CONCLUSION

This study is worthwhile and valuable in learning analytics and educational data mining. 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. Great precautions needs to be taken to protect student's privacy when trying to provide good education and services to them. Mining social media data is helpful to researchers in Analysis of student's learning Experiences.

FUTURE WORK

Apart from the work done towards this system, future work mainly comprises of the following objectives:

I. In future we can collect large student generated data other than texts which may include videos and images for analyzing the student experience with exact results and Graphs are used to show the both positive and negative experience results on a yearly basis.

II. Future work may include analysis on both positive and negative experiences as tradeoff as this category will bring essential information as well.

ACKNOWLEDGEMENT

It gives me an immense pleasure to thank many people who have supported me in completing this paper work. Firstly I would like to thank my Co-Guide, Prof. Jamgekar R. S and my guide Prof. Solunke B. R for their valuable support during the entire work. Both of them highlighted the key areas in this topic that helped me to get the right content in the topic. I am thankful to our HOD, Principle, Teacher's and college for providing us with the facilities and their guidance.

REFFERENCES

[1] Xin Chen, Student Member, IEEE, Mihaela Vorvoreanu, Krishna Madhavan,"Mining Social Media Data for Understanding Students' Learning experiences," IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 7, NO.3,JULY-SEPTEMBER 2014.

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

[3] C.-C. Chang and C.-J. Lin, "LIBSVM– ALibrary for Support Vector Machines," http://www.csie.ntu.edu.tw/_cjlin/libsvm/, 2012.

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

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

[6] G. Tsoumakas, I. Katakis, and I. Vlahavas, "Mining Multi-Label Data," Data Mining and Knowledge Discovery Handbook, pp. 667- 685, Springer, 2010.

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

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

[9] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge Univ. Press, 2000.

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

LINKS

[1] https://uberpython.wordpress.com/2012/01/01/precision-Recall-sensitivity-andspecificity.

[2] http://www.socialmediaexaminer.com/5-new-studies-show-facebook-a-marketingpowerhouse.

★★★