Understanding Students' Learning Experience by Data mining of Social Media

Ms. Pranjali S. Jadhav[1] Prof. Dr. S. S. Sane[2]

[1] ME Student, Department of Computer Engineering, *KKWIEER Nasik, Maharashtra, India*
[2] *HOD*, Department of Computer Engineering, *KKWIEER Nasik, Maharashtra, India*

ABSTRACT

The computer and Education aims to increase knowledge and understanding of way in which digital technology can enhance education. As we know, The Social Media in 21th century is very popular the Twitter and Facebook are widely used. They shared their educational experience such as their ideas, opinions, their feelings and new things about learning process. This type of knowledge (of information} is analyze and such variety of data analyzing is difficult. The students' quality expertise is mirrored in social media sites which needs human interpretation. During this paper, we have a tendency to implement 2 things analysis and huge scale data processing. I have tendency to provide example as engineering students' Twitter posts to grasp drawback occurred in their academic expertise. initial we have a tendency to get analysis and given sample taken from concerning twenty five,000 tweets associated with engineering students school life like significant study load, the social engagements etc. looking on the results, I have a tendency to apply one rule multi-label classification to classify Tweets of Students' issues.

1. INTRODUCTION

The emerging field of learning analytics and educational data mining has focused on analyzing structured data obtained from course management systems (CMS), classroom technology usage, or controlled online learning environments to inform educational decision making. However, to the best of our knowledge, there is no research found to directly mine and analyze student- posted content from uncontrolled spaces on the social web with the clear goal of understanding students' learning experiences. The drawbacks are in our study, through a qualitative content analysis, we found that engineering students are largely struggling with the heavy study load, and are not able to manage it successfully, difficulties and other psychological and physical health problems. This paper is only the first step towards revealing actionable insights from student-generated content on social media in order to improve education quality Heavy study load leads to many consequences including lack of social engagement, sleep problems, and other psychological and physical health problems. We extend the proposed algorithm which analysis the student's learning experiences by giving solutions to their problems. The suggested solution is forwarded to the student's individual email-ids to attain the privacy of student and for improving security a novel secure algorithm called BIRCH is proposed. Finally we get the feedback from the students about solution provided and comparison graph is generated.

2. LITERATURE SURVEY

Phil Long & george Siemens,[1] discussed about Analytical Learning at the Course Level and later at Departmental level. Mattias Rost, Louise Barkhuus, Henriette Cramer, Barry Brown [2], tells us about Automatic algorithm is used to get in depth meaning of FourSquare users. Mia Clark, Sheri Sheppard, Cindy Atman & others [3], give details about Academic Pathway Study Methods: Surveys, Interviews, Focus Groups, and Class room activities. Erving Goffman[4], tell us How to get best response from person by knowing previous history of his/her? Joan Morris, David R. Millen [5], explains us about Managing Personal Identity. M. Vorvoreanu, Clark, Boisvenue [6], Teach students' social media literacy. S. J. Powell, Conor Linehan, Laura Daley, Andrew Garbett, Shaun, Lawson[7], explains us about Linguistic Content Analysis. B. Pang, L. Lee, and S. Vaithyanathan[8], give us information about Support Vector Machine (SVM). Using the Twitter Search API | Twitter Developers [9], we got Application Programming Interface (API). Y.-C. Wang, M. Burke, and R. E. Kraut[10], give detail information about LDA – (Latent Dirichlet Allocation) Modeling algorithm that can detect general topics from very large scale.

3. SYSTEM FLOW

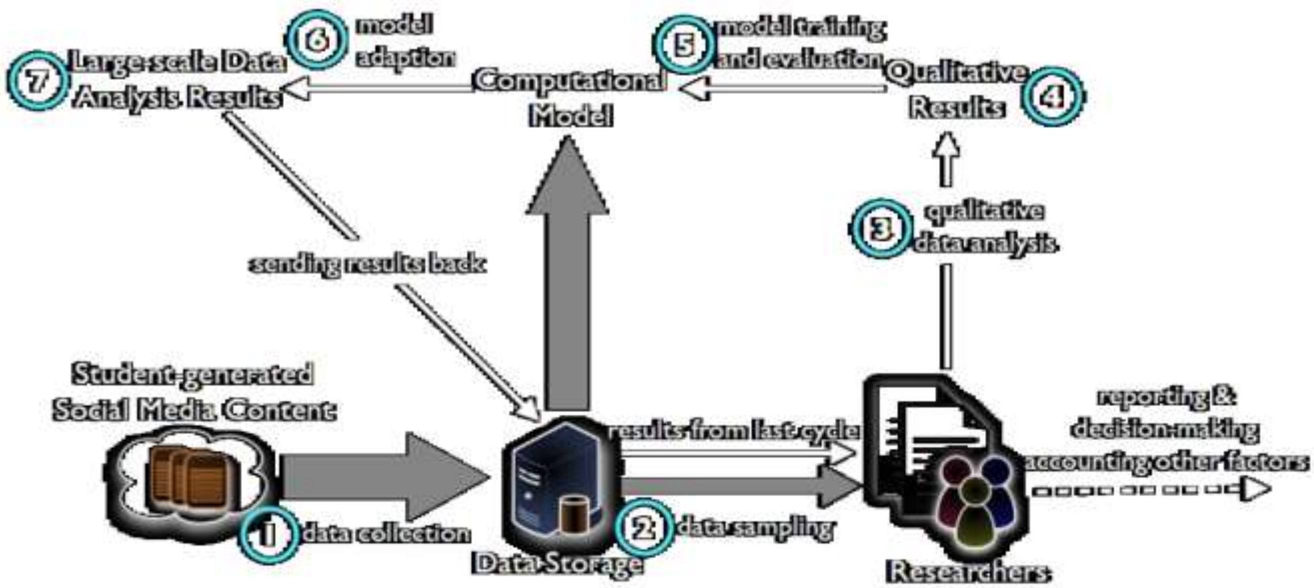

Fig.1 System Flow

3.1 Login

In this module, the user login to the social website. She/he can see own posts and the posts, posted by the engineering students.

3.2 Data collection

Collects all the data from the various students announce their comments in social web site twitter. In this, we have a tendency to conjointly collect students email-id's to send the suggestions to their individual id's.

3.3 Data clustering

In this module, the data is clustered by victimization cluster algorithmic rule. This algorithmic rule starts with single cluster. Each purpose in a info is a cluster. Then it teams nearest points into separate clusters, and continues till only 1 cluster remains. The computation of clusters calculated with facilitate of distance matrix. The algorithmic rule generates cluster feature tree whereas scanning the dataset. Every entry within the CF tree represents the cluster of objects and is characterized by triple (N, LS, SS).

3.3.1 Clustering Algorithm.

Clustering may be a method that partitions a given information set into unvaried teams supported given options specified similar objects are kept in a cluster whereas dissimilar objects are in totally different teams. It is the most necessary unattended learning drawback. It deals with finding structure in an assortment of untagged information. For higher understanding please visit Fig 2.

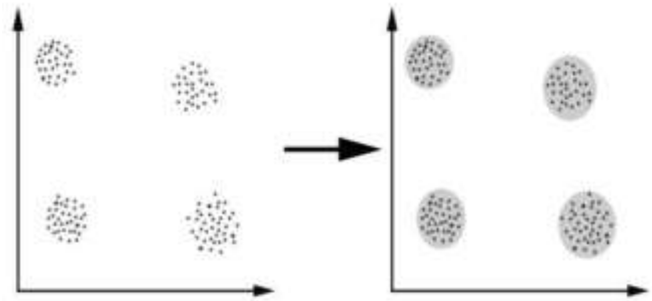

Fig. 2 Showing four clusters fashioned from the set of unlabelled data for cluster algorithmic program to be advantageous and useful a number of the conditions have to be compelled to be satisfied

3.4 Data classification

After clustering the info in several clusters supported the content, we tend to use Naïve Thomas Bayes classification algorithmic rule. One well-liked thanks to implement multi-label classifier is to rework the multi-label classification drawback into multiple single-label classification issues. One easy transformation methodology is named one-versus-all or binary relevancy. The essential conception is to assume independence among classes, and train a binary classifier for every class. Every kind of binary classifier are often remodeled to multi-label classifier mistreatment the one-versus-all heuristic.

3.4.1 Navie Bayes Classifier Algorithm:

This algorithm considers each sub words in the review and accordingly classifies the reviews in different categories:
Let S is the Sentence
Step 1: Define categories c= {c1, c2, c3,...,cn}
Step 2: Read data from a database.
Step 3: Divide S into sub works {w1, w2, w3…wn} split.
Step 4: Check sub words {w1, w2, w3…..wn} for every categories
Step 5: if words match with categories {c1, c2.c3….cn} increment the counter for that categories Else put that in "other" categories.
Step 6: Find probability of each category

3.5 Suggestions and feedback

After classification, finally we tend to send the suggestions against their issues to their individual email-id's in order that we offer privacy to the scholars and additionally get feedback from the scholars within which however useful our suggestions to them. As yet Researchers have planned a variety of techniques to handle the difficulty of SQL injection attacks. This method varies from developing best writing practices to completely machine-controlled frameworks for detection and preventing SQLIAs. This section summarizes a number of these techniques with their options and limitations.

4. DATASET

Following are the datasets for Learning Analytics and educational data mining:

4.1 Twitter post

In relationship mining, the goal is to discover relationships between variables in a data set with a large number of variables. Broadly, there are four types of relationship mining: association rule mining, correlation mining, sequential pattern mining, and causal data mining.

[Online] Available: https://dev.twitter.com/docs/using-search

4.2 Detect Students Problems from Perdue Dataset

Naïve Bayes multi-label classifier is used to detect engineering student problems from the Purdue dataset.

5. CONCLUSIONS

This information is beneficial to researchers in learning analytics, educational data mining, and learning technologies. It provides a work-flow 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. This paper informs educational administrators, practitioners and other relevant decision makers to gain further understanding of engineering students' college experiences. As an initial attempt to instrument the uncontrolled social media space, I propose many possible directions for future work for researchers who are interested in this area, good education and services to them. In the future, this analyses the student's learning experiences by giving solutions to their problems. The suggested solution is forwarded to the student's individual email-ids to attain the privacy of student and for improving security by a novel secure algorithm.

6. REFERENCES

[1] G. Siemens and P. Long, "Penetrating the Fog: Analytics in Learning and Education," *Educause Rev., vol. 46, no. 5, pp. 30-32, 2011.*
[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, 2013.*
[3] M. Clark, S. Sheppard, C. Atman, L. Fleming, R. Miller, R. Stevens, R. Streveler, and K. Smith, "Academic Pathways Study: Processes and Realities," *Proc. Am. Soc. Eng. Education Ann. Conf. Exposition, 2008.*
[4] E. Goffman, "The Presentation of Self in Everyday Life." *Lightning Source Inc., 1959.*
[5] J.M. DiMicco and D.R. Millen, "Identity Management: Multiple Presentations of Self in Facebook," *Proc. the Int'l ACM Conf. Supporting Group Work, pp. 383-386, 2007.*
[6] M. Vorvoreanu, Q.M. Clark, and G.A. Boisvenue, "Online Identity Management Literacy for Engineering and Technology Students," *J. Online Eng. Education, vol. 3, article 1, 2012.*
[7] S. J. Powell, Conor Linehan, Laura Daley, Andrew Garbett, Shaun, Lawson, "I can't get no sleep: Discussing #insomnia on Twitter", *Proc. of the 2012 ACM annual conference on Human Factors in Computing Systems*, 2012.
[8] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proc. of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, Stroudsburg, PA, USA, 2002.
[9] "Using the Twitter Search API | Twitter Developers." [Online] Available: https://dev.twitter.com/docs/using-search. [Accessed: 11-May-2013].
[10] Y.-C. Wang, M. Burke, and R. E. Kraut, "Gender, topic, and audience response: an analysis of user-generated content on facebook," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013.