ISSN 2395-1621

Student's Learning Experiences Using Data Mining

#1Supriya supekar, #2Rahul sutar, #3Reshma waje, #4S.v.ingale

#1234 Department of computer engg Bscoer, pune

ABSTRACT

Data mining is extracting meaningful data from large amount of data in this paper data mining is used to study the data for understanding students learning experiences. We develop a workflow to integrate both qualitative analysis and large scale data mining techniques. In this paper data mining techniques are used for making a decision making system. This system will used to know the problems by analyzing the status on social media posted by students. We are using multilabel classifier to classify the data which reflects problems of students facing in their educational life.This work can also prove how the informal data by social media can be useful.

Keywords— Data Mining, Classifier,Qualitative analysis

ARTICLE INFO

Article History

Received :5th April 2016

Received in revised form :

7th April 2016

Accepted : 9th April 2016

Published online :

11th April 2016

I. INTRODUCTION

Social media as twitter facebook and YouTube provide great venues for students to share joy and struggle, vent emotion and stress, and seek social support. On various social media sites, students discuss and share their everyday encounters in an informal and casual 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.

Traditionally, educational researchers have been using methods such as surveys, interviews, focus groups, classroom activities to collect data related to students' learning experiences.These methods are usually very time consuming, thus cannot be duplicated or repeated with high frequency. The scale of such studies is also usually limited.

The goal of this work are

1.To demonstrate a workflow of social media data sense-making for educational purposes, integrating both qualitative analysis and large-scale data mining techniques

2.To explore engineering students' informal conversations on Twitter, in order to understand issues and problems students encounter in their learning experiences.

II. RELATED WORK

Data Collection:

Data collection is main task in this project work. It is challenging to collect social media data related to students' experiences because of the irregularity and diversity of the language used. We started by searching based on different Boolean combinations of possible keywords such as engineer, students, campus, class, homework, professor, and lab.From this study we also get to know that most of the students uses hash tag for posting status. Students used the hash tag #engineering Problems to post about their experiences of being engineering majors. This was the most popular hash tag specific to engineering students' college life based on the data retrieved using the Boolean terms. These hash tags can also be used to retrieve data relevant to college students' experiences.

Inductive Content Analysis

Because social media content like tweets contain a large amount of informal language, sarcasm, acronyms, and misspellings, meaning is often ambiguous and subject to human interpretation.

There were no pre-defined categories of the data, so we needed to explore what students were saying in the tweets. Thus, we 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.

Development Of Categories:

There are five categories in which we are going to classify our data.The five categories are Heavy study load, Lack of social engagement, Negative emotions, Sleep problems ,diversity issues.The comments or the data we are using is going to be classify in this five categories by which we will get to know what problems students are facing.

III.SYSTEM ARCHITECTURE

System Architecture:-

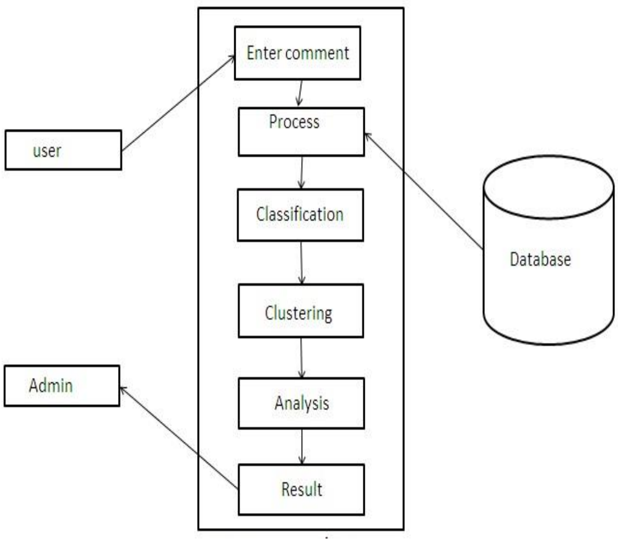

Figure 1: System Architecture diagram

This is the system architecture design. In this, first the user will enter the comment.The comment will be stored in database. Again the data from Database will be preprocessed.Then classification will be done on that data by using Naïve Bayes Algorithm..After that clustering process will done.After clustering process data will be analysed.And at the end the result of the analysis is shown to admin.
This system provides a more user friendly functionalities and attractive user interface, some of the functionalities are listed below:

- Connetivity between social media,
- Collecting data from social media,
- Identifying keywords,
- Classification Algorithm
- Clustering the data.

Data Clustering

In this module, the raw data is clustered by using clustering algorithm. This algorithm starts with single cluster. Every point in a database is a cluster. Then it groups closest points into separate clusters, and continues until only one cluster remains. The computation of clusters calculated with help of distance matrix. The algorithm generates cluster feature tree while scanning the dataset. Each entry in the CF tree represents the cluster of objects and is characterized by triple (N, LS, SS).

Data Classification

After clustering the data in different clusters based on the content, we use Naïve Bayes classification algorithm. One popular way to implement multilabel classifier is to transform the multilabel classification problem into multiple singlelabel classification problems. One simple transformation method is called one versus all or binary relevance. The basic concept is to assume independence among categories, and train a binary classifier for each category. All kinds of binary classifier can be transformed to multilabel classifier using the one versus all heuristic.

Suggestions And Feedback:

After classification, finally we send the suggestions against their problems to their individual email id’s so that we provide privacy to the students and also get feedback from the students in which how helpful our suggestions to them.

IV.RESULTS

Result will be the classification of the data or comments in predefined categories.Some of that data may classify into Heavy study load, Lack of social engagement, Negative emotions, Sleep problems ,diversity issues.It will be calculated by using naïve bayes classifier.

V. FUTURE SCOPE

This tool will provide a friendly user interface and integration between qualitative analysis and the classification and detection algorithms. Therefore, educators and researchers using this tool can focus on the actual data analysis and investigate the types of learning issues that they perceive as critical to their institutions and students. This tool can also facilitate collaboration among researchers and educators on data analysis. Advanced natural language processing techniques can be applied in the future to provide topic recommendations and further augment the human analysis results, but cannot completely rule out the human effort.

VI. CONCLUSION

Our study is beneficial to researchers in learning analytics, educational data mining, and learning technologies. 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. Our study can inform educational admin- illustrators, practitioners and other relevant decision makers to gain further understanding of engineering students college experiences.

REFERENCE

[1].M. Clark, S. Sheppard, C. Atman, L. Fleming, R. Miller, R. Stevens, R. Streveler, and K. Smith, “Academic pathways study: Processes and realities,” in Proceedings of the American Society for Engineering Education Annual Conference and Exposition, 2008.

[2]. R. Baker and K. Yacef, “The state of educational data mining in 2009: A review and future visions,” Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2009.

[3]. G. Siemens and P. Long, “Penetrating the fog: Analytics in learning and education,” Educause Review, vol. 46, no. 5, pp. 30– 32, 2011.

[4]. R. Ferguson, “The state of learning analytics in 2012: A review and future challenges,” Knowledge Media Institute, Technical Report KMI-2012-01, 2012.

[5]. M. Rost, L. Barkhuus, H. Cramer, and B. Brown, “Representation and communication: challenges in interpreting large social media datasets,” in Proceedings of the 2013 conference on Computer supported cooperative work, 2013, pp. 357–362.

