User Domain Analysis Using Social Media

Prof. Rinku A. Badgujar[1], Pranjali Maske[2], Anju Patil[3], Mahananda Tidke[4]
[1, 2, 3, 4] Department of Computer Science
[1, 2, 3, 4] JSPM(BSIOTR) Wagholi, Pune, India

Abstract- *Users informal conversation on social network can be used to know their educational experiences such as opinions, concerns and learning process. These data from such sites will help us to know about users problem and their difficulties. The complexity is that we need human interpretation in analysis of the data. In this paper, we developed a workflow to integrate both successful analysis and big-scale data mining techniques. We focused on people posts to understand issues and problems in their experiences. We then used the algorithm to train a detector of users problem from about 40,000tweets. This presents a methodology and results that show how common communication on social media data can provide insights into users analysis.*

Keywords- Social networking, web-text analysis, Education, Social network analysis, Computer and Education, Data mining.

I. INTRODUCTION

The Educational Institutions have been traditionally using methods such as surveys, interviews, focus groups, classroom activities to collect data user's experiences. There are so many traditional methods available such as questionnaires, surveys and interviews to analyze the users barriers in an educational institution. But the main problem with these methods is these techniques are require more time and cannot be performed efficiently with large frequency as the analysis has to be performed randomly

So there is a need for analysis of social media data qualitatively with the help of integrating both the flow and algorithm approach. As one of the largest social network on internet facebook is helpful for expanding in your business. This is gives different advice to user or developers, who want to get started. Many people are able to share their ideas on social media for business is beneficial, it is very effective for marketing. Therefore, this mechanism has failed in conveying users concern, as desired by the educational decision-makers & researchers. We propose automation system in extracting and mining data, through the informal posts and chats on social media platforms, made by the users, in order to exactly know about their concerns and issues, on a larger scale.

As it has been seen, across several social media platforms, users are informally or casually posting about their concerns and feelings on the social media platforms and refrain from making such feelings or concerns available through feedbacks or surveys employed by the educational systems. In this system, the user's 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 support vector Machine(SVM) for classification to demonstrate the workflow of social media data sense-making for research, education purposes, integrating both qualitative analysis & large scale data mining techniques. Proposed system developed a flow to put together both successful investigation & big scale data mining techniques. it paying attention on users posts to know problem & trouble in their social.

II. SCOPE

Scope of the system:

1) Read all the Post present on the Social Media and Analysis of that Post.
2) Graph generation on the classified Post.
3) Remove the error full or noisy data.
4) Applying Support vector Algorithm For Stopword and stremming data.
5) Graph is generated as per the different domain's post analysis.

III. SYSTEM ARCHITECTURE

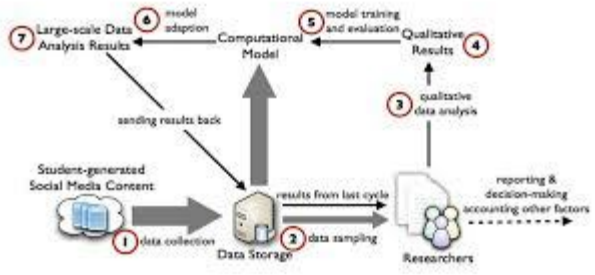

Fig. System Architecture

Step 1: Data Classification In this stage an inductive content analysis on samples or procedures of the learning problems data set. This type of procedure is called as data sampling.

Step 2: Qualitative Analysis of data in this step of qualitative data analysis it identifies the categories.

Step 3: Qualitative Analysis of data in this step of qualitative data analysis it identifies the categories.

Step 4: Qualitative Results of data As per the step 3 of qualitative data analysis it identifies the categories and then in step 4 it returns the keywords as per the categories

Step 5: Evaluation of model training as per the categories we can implement a ideas based classifier to analyze and classify the posts.

Step 6: In this step the classification algorithm to train a detector that could assist detection of higher educational users problems.

Step 7: Large-Scale Analysis Results of data this is the last step of the process, it returns the total final results in the classified format with the graphical analysis.

IV. ALGORITHM

SVM Algorithm:

- Process for mining social media data
  1. Train the sample vector
  2. Stages for processing data

- Gathering the training data
- Pre-processing the data and remove the noisy and unwanted data, by using a particular earlier trained vectors.
- Remove the Stop words. e.g."a","the", prepositions "in", "of", "by" and some other short words like n't, doesn't .
- Use Stemming for classifying the data is noisy data.
- Gather the text from the pre-processed data, and store it for processing.
- We use the earlier trained sample data set, to differentiate the data into categories. Based upon the feature vector of the trained vectors, such as the weight of keywords, we differentiate the post into their respective category.

Feature analysis:

- We determine the post category based on below steps:

1. We find out the weight of keywords from the post that matches the keywords from the training data set.
2. Later on we find out the percentage of the matching

$$\text{particular category} = \frac{(\text{no of keys of a particular training set})}{(\text{total no of keys found in the post})} * 100$$

3. Later on we give the keyword weight as input to each category SVM to identify similar categories.
4. out of the selected categories identify category having maximum percentage and add post to that category.

V. RESULT

Registration Page -

A user is created in the system by taking into consideration certain essential data fields. A user can log in to the system by inserting his credentials.

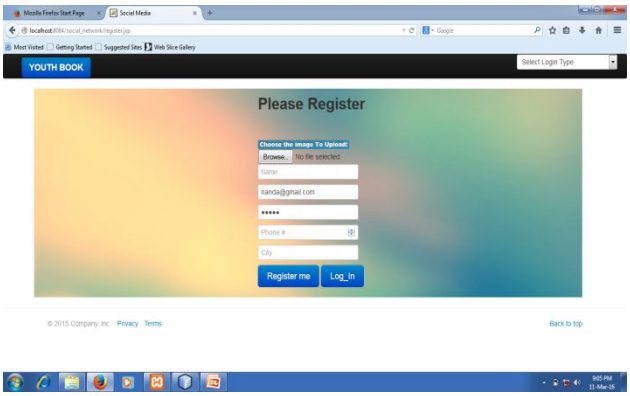

Fig..Registration Page

Login Page –

Fig. Login Page

First Run the Login form showing username and password. Enter the legal username and password. Select the Authentication button after entering username and password.

Add And View Posts-

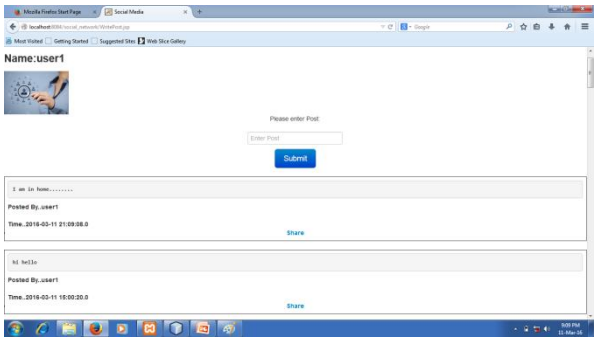

Fig. Add And View Posts

The user has the rights to publish his views in the form of a post in the application

View All Users Posts-

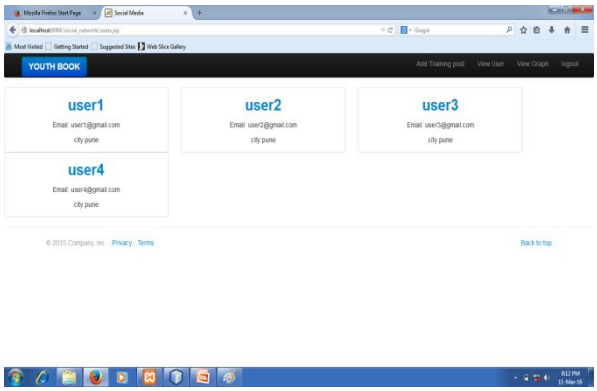

Fig. View all users posts

View Graph of Result-

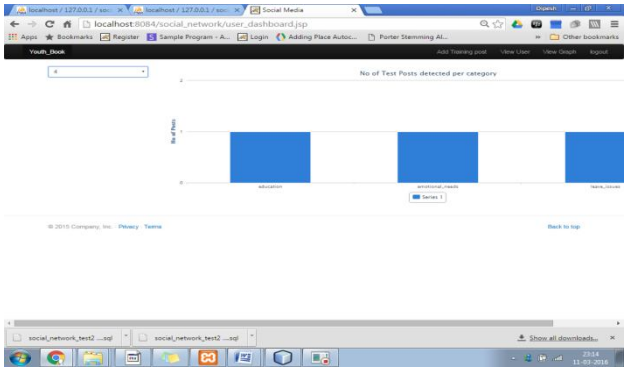

Fig. View graph of result

Identify post and categories. From post stop word are removed and streaming is done. Remove stream words, TFIDF(Term Frequency and Inverse Document Frequency) Calculations and cosine similarity of every keywords in post are calculated. support vector machine analysis and training part is done by the admin side. Finally large scale graph are generated.

VI. ACKNOWLEDMENT

The authors wish to thank Xin Chen for their contributions to this study. This work was supported in part by JSPM's Institute Pune.

VI .CONCLUSION

In our study is beneficial to researchers in learning analytic, educational data mining, and learning technologies. It provides a flow for analyzing social media data for education, marketing etc purposes that overcomes the large disadvantages of both successful analysis and big scale computational analysis of user-generated textual content. Our study can inform educational people, infrastructure and other relevant decision makers to gain further understanding of users experiences. As an initial attempt we propose many possible direction for future work for researchers who are interested in this area.

REFERENCES

[1] Xin Chen, MihaelaVorvoreanu, and Krishna Madhavan.” Mining Social Media Data for Understanding Students’ Learning Experiences” IEEE transactions on learning Technologies, ID, DOI 10.1109/TLT.2013.2296520

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

[3] 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.

[4] 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.

[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] Rinku Badgujar, "An Efficient Data Anonymization Technique-Slicing", Journal Of ISSN, vol. 1, no. 8,2014.

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

[8] E. Goffman The Presentation of Self in Everyday Life, 1959 :Lightning Source Inc.

[9] C. Moller-Wong and A. Eide "An Engineering Student Retention Study", J. Eng. Education, vol. 86, no. 1, pp.7 -15 1997

[10] National Academy of Eng. The Engineer of 2020: Visions of Engineering in the New Century, 2004 :National Academies Press