Survey on Efficient Dynamic Induction Model for Mining Social Media Data with Different Mining Methods

Goodlesh Dugad, Prof. Sharmishta Desai

M.E. Student, Dept. of Computer Engineering, Savitribai Phule Pune University, Pune, India [1]

Professor, Dept. of Computer Engineering, Savitribai Phule Pune University, Pune, India [2]

ABSTRACT: In today online world peoples are dependent on a social media for communication as well as expressing the feelings and information with the friends/peoples. Social media became a platform on which people feels free to express their emotions. The most used users of social media are the students. Students' social media behavior uncovers about their everyday life. Students post their experiences on social networking sites be it individual or academic. Analysis of these posts, in any case, is not a simple job. Pure manual analysis is not productive as information increments at a quick rate. There is a work process created by survey topic that acclimatizes the qualitative analysis as well as data mining methods. The attention is principally on engineering students' posts so their issues can be investigated. This makes use of the approach of multi-label classification, which is because of working of classes among students' problems-heavy study load, lack of sleep, absence of social engagement, and so on. This empowers web-based social networking to reveal insight into students' academic experiences. This survey presents a study of different methods performed by different researchers on the social media mining for student learning experiences. Also dynamic induction model is proposed for mining this social media data. Proposed model is implemented using twitter data set. As data is big in size our proposed method will perform well as compared to traditional methods like nave bayes.

KEYWORDS: Energy efficient algorithm; Manets; total transmission energy; maximum number of hops; network lifetime

I. INTRODUCTION

Data mining is a powerful technology for recognizing useful patterns in complex data. It has repeatedly provided proven results in information systems. Its use has benefited many sectors, such as banking, retail, marketing, biology, medicine, telecommunication, and others, resulting in significant advancements for these industries.

Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Operationally, automated processes come together to extract useful information— patterns, associations, changes, trends, anomalies, and significant structures—from large or complex data sets. Educational data mining is the process of collecting and analyzing a wide range of student data in order to derive knowledge about learning habits and behaviors in order to personalize educational interventions to maximize student outcomes.

 Data present in social media are unstructured. These online networking gives such a large number of helpful learning and data about the students' emotions, feelings, experiences and struggles in their reviews outside the classroom. So taking after these Students track on the social media and it's an interesting outlook for instructive analysts and professionals to comprehend student learning experiences outside the classroom. This analysis will reveal such a large number of experiences which was not cleared or considered while the classroom discourses. This study about their experiences gives so helpful information to the improvement of students in decision making, enhancing students' education quality, training and placement, withholding and achievements.

The measure of social media information gives opportunities to comprehend students' experiences, yet their methodological challenges to utilize web-based social networking information for educational reason. In classroom

studies, to know every single students' perspective overviews, surveys, group discussions, meeting and hiding was finished. The inventive idea of utilizing online networking information concentrates on the required data and knowledge to be extricated for educational reason by comprehension students' experiences.

Many of mining methods for the purpose of the text document are utilized for mine different textual matches via social media. Specific social sites application with social websites such as Facebook, Twitter, and Youtube provides numerous options to ascertain communication via people leading to communal learning and circulation of important knowledge such as comments, posts, emotions, along with likes.

II. LITERATURE REVIEW

In paper [1] authors implemented a system which gives workflow to integrate both qualitative analysis and large-scale data mining techniques. By focusing on engineering students' twitter posts to understand issues and problems in their educational experiences. They found engineering students encounter problems such as heavy study load, lack of social engagement, and sleep deprivation. Based on several test results conducted by authors, they implemented a multi-label classification algorithm to classify tweets reflecting students' problems.

In paper [2] authors proposed partial supervised learning for HDP which enables HDP to make use of partial known knowledge to guide the model learning process. They applied the proposed partial supervised learning for HDP to classify posts (micro-blogs) in an educational environment.

In paper [3] authors proposes a novel application of text categorization to identify relevant and irrelevant micro-blogging questions asked in a classroom. Several modeling approaches and several weighting or preprocessing configurations are studied for this application through extensive experiments.

In paper [4] authors e propose a two-step analysis framework that focuses on positive and negative sentiment, as well as the side effects of treatment, in users' forum posts, and identifies user communities and influential users for the purpose of ascertaining user opinion of cancer treatment. After analysis authors introduced a novel network-based approach for modeling users' forum interactions and employed a network partitioning method based on optimizing a stability quality measure.

In paper [5] authors explored privacy concerns related to mining social media networks. Specifically, author looked at the issue through a crime incident mining context, looking at matters related to social media data ownership, legal protection of personal information, methods that may be used to anonyms users.

In paper [6] authors implemented novel data-mining method was developed to gauge the experiences of medical devices and drugs by patients with diabetes mellitus. Self-organizing maps were used to analyze forum posts numerically to better understand user opinion of medical devices and drugs. The implication of this novel data-mining method could open new avenues of research into rapid data collection, feedback, and analysis that would enable improved outcomes and solutions for public health.
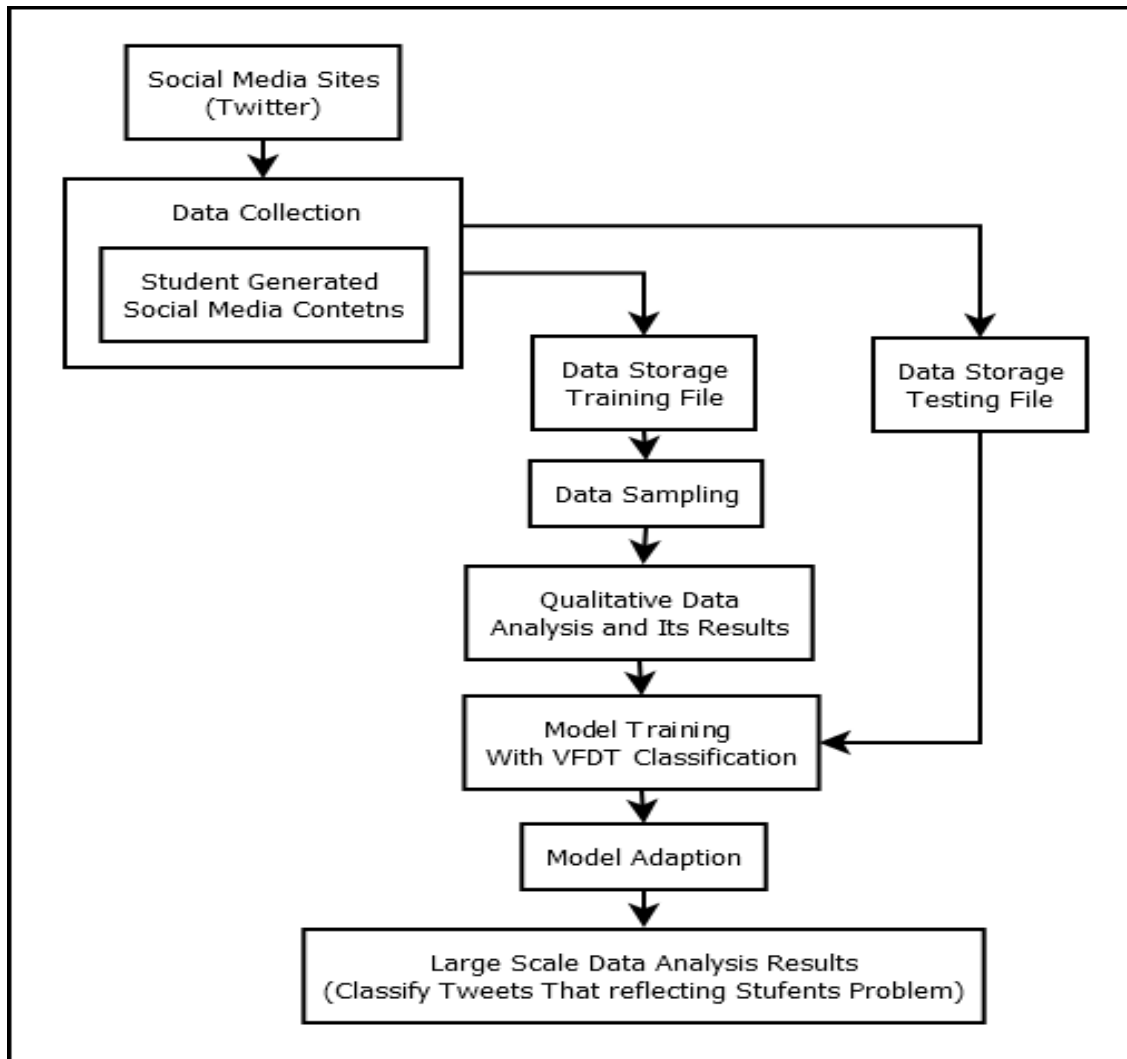
III. PROPOSED SYSTEM

Fig. System Architecture

Figure shows the proposed system architecture. Data from the students post on twitter are collected in database; extraction is done on the data. Results are then collected and noisy data is removed. Refining the data from the store gives the model training evaluation. Finally by the model adaption, large scale data analysis result is generated. First system collects posts by engineering students on the social networking site. The tweets are stored in database and inductive content analysis is then performed on the data storage training file. We categorize their several problems in prominent categorizes. Pre-processing of the data's is done. Based on these categories, we are implementing a VFDT classification algorithm. Thus the evaluation results could help educators to identify at-risk students and make decisions on proper interventions to retain them.

IV. CONCLUSION

This paper analyses various techniques used for data mining from social media. Also given the advantages and drawbacks present in the different studies performed by various researchers. To deal with drawbacks in present systems we presented an idea of the new system.

REFERENCES

1. X. Chen, M. Vorvoreanu and K. Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences," quot; in IEEE Transactions on Learning Technologies, vol. 7, no. 3, pp. 246-259, July-Sept. 2014.
2. D. Wang, A. Al-Rubaie, A. A. Dhanhani and J. Ng, "Smart text-classification of user-generated data in educational social networks," Frontiers in Education Conference (FIE), 2015. 32614 2015. IEEE, El Paso,TX, 2015, pp. 1-5.
3. S. Cetintas, L. Si, H. P. Aagard, K. Bowen and M. Cordova-Sanchez, and quot;Microblogging in a Classroom: Classifying Students-39; Relevant and Irrelevant Questions in a Microblogging-Supported Classroom,; inIEEE Transactions on Learning Technologies, vol. 4, no. 4, pp. 292-300, Oct.-Dec. 2011.
4. A. Akay, A. Dragomir and B. E. Erlandsson, ";Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care", in IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 1,pp. 210-218, Jan. 2015
5. N. Moorosi and V. Marivate, "Privacy in mining crime data from social Media: A South African perspective," 2015 Second International Conference on Information Security and Cyber Forensics(InfoSec), Cape Town, 2015, pp. 171-175.
6. A. Akay, A. Dragomir and B. E. Erlandsson, "A novel data-mining approach leveraging social media to monitor and respond to outcomes of diabetes drugs and treatment", 2013 IEEE Point-of- Care HealthcareTechnologies (PHT), Bangalore, 2013, pp. 264-266.
7. D. Konopnicki et al., "A statistical approach to mining customers"39; conversational data from social media", in IBM Journal of Research and Development, vol. 57, no. 3/4, pp. 14:1-14:13, May-July 2013.
8. D.M. Romero, B. Meeder, and J. Kleinberg, "Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Politi- cal Hashtags, and Complex Contagion on Twitter," Proc. 20th Int'l Conf. World Wide Web , pp. 695-704, 2011.
9. www.carnegiefoundation.org/blog/the-learning-analytics-landscapetension-between-student-privacy-and-the-process-of-data-mining/
10. Dr Nirmala C R, Roopa G M, Naveen Kumar K R "Twitter Data Analysis for Unemployment Crisis"IEEE conference 2015