Sentiment Analysis for Understanding Students' Learning Experiences: A Survey Paper

ME Scholar Aaquib Multani   **Assistant Professor Atul Agrawal**
Atul.273@gmail.com   multaniaaquib@gmail.com
Dept. of Computer Science
Sushila Devi Bansal College of Engineering
Indore, MP., India.

Abstract- Social media allows the creation and interactions of user created content. Social medium places include Facebook, Twitter etc. Student's casual discussion on social media focused into their educational experience, mind-set, and worry about the learning procedure. Now days, Social media craze is mounting to heights of success for every individuals. The amount of social media data provides chances to understand students' experiences, but their methodological difficulties to use social media data for educational purpose. Therefore, there is need to identify students' emotions by analysis of their sentiments. In this survey paper, we discuss about the sentiments analysis and their patterns. Additionally, we proposed an approach for predicting sentiments of user specifics and classifying them in to 'negative' or 'positive'. We are implementing this method on JAVA environments using "Twitter" dataset.

Keywords- Sentiment Analysis, Social Media, Twitter, Dataset, Text Mining, Decision Tree, Naïve Bayes.

I. INTRODUCTION

Social media sites such as Twitter, Face book, and YouTube provide great venues for students to share joy,[1] and struggle, vent emotion and stress, and seek social support Student's discuss and share their everyday encounters in formal and informal manner on different social media sites. These 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. 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.

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. 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 [3]. We propose to address this problem by creating a system that analyses students' learning experiences in real-time. To create such as system, sentiment analysis can be used.

Sentiment analysis is an application of natural language processing, computational linguistics and text analytics that identifies and retrieves sentiment polarity from the text by studying the opinion. Sentiment polarity is usually either positive or negative, although sometimes neutral is included [10]. 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:

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

II. LITRATURE SURVEY

This section provides the recently made contribution and the research work performed for improving Learning Experience using sentiment analysis technique. Thus different research articles and papers are included in this section. In the paper of Mining Social Media Data for Understanding Students Learning Experiences, they have developed a workflow to integrate both qualitative analysis and large scale data mining techniques.

They have focused on engineering students' twitter posts to understand issues and problems in their educational experiences. They have first conducted a qualitative analysis on samples taken from about 25,000 tweets related to engineering students' college life [1].In this

paper, they found that engineering students encounter problems such as heavy study loads, lack of social engagement, and sleep deprivation. Based on this result, they implemented a multi-label classification algorithm to classify tweets reflecting students' problems. In the paper of Sentiment Analysis of Tweets using SVM analyzed the performance of Support Vector Machine (SVM) for sentiment analysis. For performance analysis of SVM, they used two pre classified datasets of tweets, first dataset consisted of tweets regarding self-driving cars and second dataset was about the apple products. Weka tool is used for performance analysis and comparison. Results are measured in terms of precision, recall and f-measure. According to results, for first dataset the average precision, recall and f-measure is 55.8%, 59.9% and 57.2% respectively.

For second dataset the average Precision, Recall and F-Measure is 70.2%, 71.2% and 69.9% respectively. Complete results are shown in tabular and in graphical forms. The results clearly show the dependency of SVM performance upon input dataset. The performance dependency of SVM and other machine learning techniques should be explored further by using large and different datasets. For comparative analysis the results of this paper can be used as baseline. Moreover it should also be investigated that for classification purpose, which machine learning algorithm performs better on which type of dataset and what might be the reasons? This can lead the researchers to the improved versions of machine learning algorithms for classification purpose, [14]

In the paper of Sentiment analysis in twitter using Natural Language Processing (NLP) and classification algorithm A live Twitter feed is collected under the keywords entered by the user. The feed is stored in a MongoDB database. It is also stored locally in a json file. The data was pre-processed to remove unnecessary spaces, symbols and useless features. It still requires further work to remove as much noise as possible. Approximately over 2000 tweets are then stored as a csv file for analysis.

A number of Lexicon based methods are utilized on individual tweets from the file to assess their usefulness. The chosen classifier for this work is a Naive Bayes Classifier utilizing the text processing tools in NLTK and their capacity to work with human language data. It is trained on tagged tweets and then used to analyze the sentiment in the tweets about the searched topic. The result is represented in the form of a pie diagram which shows the percentage of users who have positive opinion on the searched topic as compared to the ones have negative opinion or are neutral[15]. In the paper of Analyzing Social Media Data in Educational Sectors Using Data Mining Techniques they 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. And the study can inform educational administrators, practitioners and other relevant decision makers to gain further understanding of engineering student's college experiences. It also advocates that great attention needs to be paid to protect students privacy when trying to provide good education and services to them [16].

They analyzed twitter data for product and market review using R package they found it is possible to find the sentiment analysis of them .They apply text mining tasks and sentiment analysis for twitter data to analyze user contributed reviews for products or services. It can be generalized that, Businesses can utilize their consumer opinions generated from social media tracking and analysis by adapting their marketing plans, products and business intelligence respectively [17].

In the paper of fake news detection on social media they detect which news is fake or real by they explored the fake news problem by reviewing existing literature in two phases: characterization and detection. In the characterization phase, we introduced the basic concepts and principles of fake news in both traditional media and social media. In the detection phase, we reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction. [7].

III. TAXONOMY
Taxonomy is the classification of terminology which is frequently used for the purpose of the task completion. Taxonomy helps us to make information available in precisely.

1. Natural language processing
Natural Language Processing (NLP) is the practical field of Computational Linguistics, although some authors use the terms almost interchangeably. Sometimes NLP has been considered a sub discipline of Artificial Intelligence, and more recently it sits at the core of Cognitive Computing, since most cognitive processes are either understood or generated as natural language utterances.NLP is a very broad topic, and includes a huge amount of sub divisions: Natural Language Understanding, Natural Language Generation, Knowledge Base building, Dialogue Management Systems (and Intelligent Tutor Systems in academic learning systems), Speech Processing, Data Mining – Text Mining – Text Analytics, and so on [12].

2. Text Mining
Text Mining is the automated process of detecting and revealing new, uncovered knowledge and inter-relationships and patterns in unstructured textual data

resources. Text mining targets un-discovered knowledge in huge amounts of text. Whereas, search engines and Information Retrieval (IR) systems have specific search target such as search query or keywords and return related documents

Firstly, a set of un-structured text documents is collected. Then, the pre-processing for the documents is performed to remove noise and commonly used words, stop words, stemming [11]. This process produces a structured representation of the documents known as Term-document matrix, in which, every column represents a document and every row represents a term occurrence throughout the document. The final step is applying data mining techniques such as clustering, classification association rules to discover term associations and patterns in the text and then, finally, visualizing these patterns using tools such as word-cloud or tag-cloud.
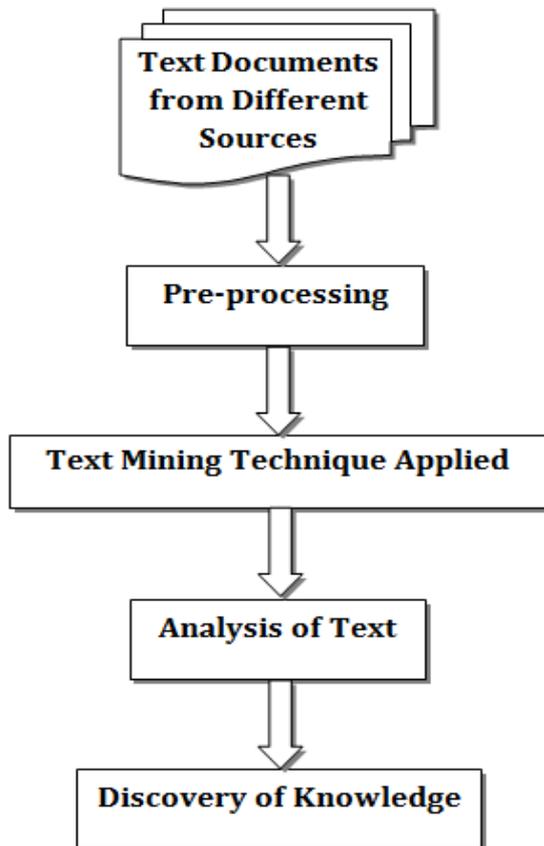

Fig. 1 Text Mining Process [11]

3. Sentiment analysis

Sentiment analysis is an application of natural language processing, computational linguistics and text analytics that identifies and retrieves sentiment polarity from the text by studying the opinion. Sentiment polarity is usually either positive or negative, although sometimes neutral is included. Previous research has shown that sentiment

analysis is more effective when applied to specific domains[13].

To the best of our knowledge, sentiment analysis has not been applied for analyzing students' learning experiences. Consequently, there is need of investigating different models and look at the best combination of pre-processing methods, features and machine learning techniques to create the best-suited model for our purpose.
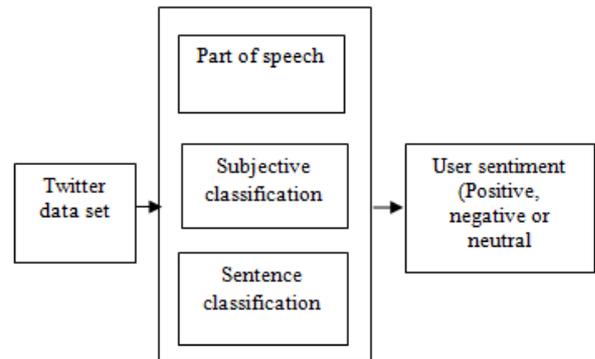

Fig 2 Sentiment Analysis Process.

4. Navie Bayes classifier

This is a simple probabilistic classifier that is based on the Bayesian probability. The Naive Bayes classifier is based the assumption that feature probabilities are independent of one another. This classification technique assumes that the any feature in the document is independent of other feature. Naive Bayes classifier considers a document as collection of words and assumes that the probability of a word in the document is independent of its position in the document and the presence of other word .We derive the Naive Bayes (NB) classifier by Bayes' rule,

$$p\left(\frac{c}{d}\right) = \frac{p(c)p(d/c)}{p(d)}$$

Where $P(d)$ plays no role in selecting c. But its conditional independence assumption clearly does not exist in real-world situations, Naive Bayes-based text classification still tends to perform well.

5. Decision Tree

According to the datasets by information extraction, a decision tree constantly updated data to update the decision tree, and then generate the understandable rules. The experiment proves that it is feasible to realize the Web information extraction based on the decision tree [2]. Decision Tree Technology, which has three courses.
•Construct wrapper.
•Decision tree building process, a rough decision tree will be constructed based on an algorithm.
•Decision tree refinement process and to automatically extracted knowledge or rules.

6. Emotion classification

Emotion classification aims to detect and recognize types of feelings through the expression of texts, such as anger, disgust, fear, happiness, sadness, and surprise. Emotion can be expressed in many ways that can be seen such as facial expression and gestures, speech and by written text. Emotion Detection in text documents is essentially a content - based classification problem involving concepts from the domains of Natural Language Processing as well as Machine Learning.

IV. PROPOSED WORK

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 paper we intend to develop data mining system using Stop-word and stemming for pre-processing, Naive Bayes and C4.5 (Decision Tree) for Classification to demonstrate the workflow of social media data sense-making for educational objective, fussing both qualitative analysis and various data mining techniques. In this paper we are going and proposed and accurate data model that enhances the sentiment analysis over the different post of students'.For the students' experiences of particular study in term of positive and negative and neutral emotions. The method of proposed model shown by block diagram
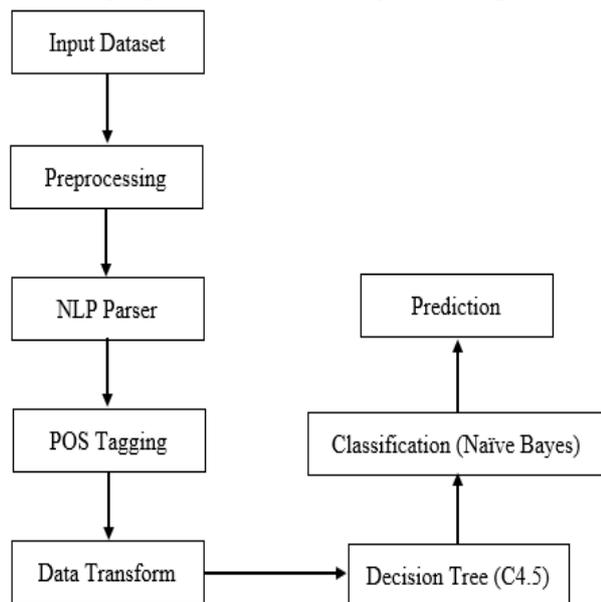

Fig. 3 Proposed System Architecture .

1. System flow- The system flow is shown in the figure below. In this system there is an investigative procedure to find the appropriate data.

- Gathering of tweets/post/data from various social media sites such as twitter. This corresponds to the step 1 In Fig. 3.
- Next in the step 2 we perform Preprocessing on the unstructured tweets. Preprocessing is the process of cleaning the data from unwanted elements. It increases the accuracy of the results by reducing errors in the data. There are many general preprocessing techniques, of which the most common are: tokenization, covert text to lower or upper case, remove punctuation, remove numbers, remove repeated letters, remove stop words, stemming and negation.
- After in the step 3 anatural language parser mainly program that works out the grammatical structure of sentences ,for instance which groups of words go together as phrases ,subjects, object or verbs.
- Part of speech: It is the process marking up text according to part of speech. It uses noun, verb, adverb, adjectives for identification of words.
- In next step we transform the data for visualization.In data transform we mainly convert data one form into other. Common examples include filtering and grouping of data.
- After that we use classifier C4.5 Decision trees are a very effective method of supervised learning. It aims is the partition of a dataset into groups as homogeneous as possible in terms of the variable to be predicted. It takes as input a set of classified data, and outputs a tree that resembles to an orientation diagram where each end node (leaf) is a decision (a class) and each non- final node (internal) represents a test. Each leaf represents the decision of belonging to a class of data verifying all tests path from the root to the leaf.
- We use another classification after decision tree which is naive bayes classificationthe performance of the classifiers is estimated by comparing it with other multi label. In the classification algorithm is applied by System to prepare 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.

V. CONCLUSION

Our study is beneficial to researchers in learning analytics, educational data mining, and learning technologies. It provides a workflow for analyzing social 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 administrators, practitioners and other relevant decision makers to gain

further understanding of engineering students' college experiences [1].

As an initial attempt to instrument the uncontrolled social media space, we propose many possible directions for future work for researchers who are interested in this area hope to see a proliferation of work in this area in the near future. We advocate that great attention needs to be paid to protect students' privacy when trying to provide good education and services to them [9].

REFERENCES

[1] Xin Chen, Mihaela and Krishna P.C, "Mining Social Media Data forUnderstanding Students Learning Experiences", IEEE Transactions on Learning Technologies, 2014.

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

[3] S. Cetin as, L. Si, H. Aagard, K. Bowen, and M. Cordova-Sanchez, "Micro blogging in Classroom: Classifying Students' Relevant and Irrelevant Questions in a Micro bloggingSupported Classroom," Learning Technologies, IEEE Transaction son, vol. 4, no. 4, pp. 292–300, 2011.

[4] Zhou, X., Tao, X., Yong, J., & Yang, Z. (2013, June). Sentiment analysis on tweets for social events. In Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on (pp. 557-562). IEEE.

[5] Danneman, N., & Heimann, R. (2014). Social Media Mining with R. Packt Publishing Ltd Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of natural language processing, 2, 627-666.

[6] Ayesha Rashid et al, "A Survey Paper: Areas, Techniques and Challenges of Opinion Mining", International Journal of Computer Science (IJCSI), Vol 10 Issue 6 No 2, Nov 2013.

[7] Fake News Detection on Social Media: A Data Mining PerspectiveKai Shuy, Amy Slivaz, Suhang Wangy, Jiliang Tang \, and Huan LiuyyComputer Science & Engineering, Arizona State University, Tempe, AZ, USAzCharles River Analytics, Cambridge, MA, USA\Computer Science & Engineering, Michigan State University, East Lansing, MI, USA 2018

[8] K. Nirmala, S. Satheesh kumar and Dr. J. Vellingiri "A Survey on Text categorization in Online Social Networks" International Journal of Emerging Technology and Advanced Engineering Volume 3, Issue 9, September 2013

[9] Survey on Social Media using Twitter Data for Students Experience using Naïve Bayes TechniqueInternational Institution for Technological Research and Development Volume 1, Issue 1, 2015

[10] A Study on Sentiment Analysis: Methods and Tools Abhishek Kaushik1, Anchal Kaushik2, Sudhanshu Naithani3International Journal of Science and Research (IJSR) ,2014

[11] Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study

[12] Text Analytics: the convergence of Big Data and Artificial Intelligence Antonio Moreno1, Teófilo Redondo2 Special Issue on Big Data & AI ,2013

[13] J. Yang and S. Counts, "Predicting the speed, scale, and rangeof information diffusion in twitter," Proc. ICWSM, 2010.

[14] Sentiment Analysis of Tweets using SVM by International Journal of Computer Applications (0975 – 8887) Volume 177 – No.5, November 2017

[15] Pranav Waykar, Kailash Wadhwani, Pooja More, "Sentiment analysis in twitter using Natural Language Processing (NLP) and classification algorithm",International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 1, January 2016

[16] Manjunath B Talawar and Prasad.M. R, Analyzing Social Media Data in Educational, Sectors Using Data Mining Techniques, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (4), 2016

[17] Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study Eman M.G. Younis International Journal of Computer Applications (0975 – 8887) Volume 112 – No. 5, February 2015.