ISSN: 2321-7782 (Online)
Impact Factor: 6.047
Volume 4, Issue 5, May 2016
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

Understanding Students' Learning Experiences Using Social Media Data Mining

Suahant Patil[1]
Pune University, Department of Computer Engineering,
IOK College of Engineering,
Pune, Maharashtra – India

Prof. Ajay Gupta[2]
Pune University, Department of Computer Engineering,
IOK College of Engineering,
Pune, Maharashtra – India

***Abstract:** Because of quick development of web user, Social Networks have ended up one of the respected correspondence medium utilized over web. A large number of messages are showing up consistently on popular sites that give web services, for example, Twitter, Facebook, and LinkedIn. A large number of users impart their own insights or perspectives about on different of issues and talk about a few current hotly debated issues on Twitter, making it an essential base for following and dissecting sentimentation of students. Such following and analysis can give basic data to basic leadership or conclusion mining in assortment of areas. In this our work, we have moved above and beyond to decipher notion variations. We watched that developing themes (named closer view subjects) inside the sentiment variation periods are exceptionally identified with the certified explanations for the variations. We select the most illustrative tweets for closer view subjects and build up another generative model called Reason Candidate and Background LDA (RCB-LDA) to rank them concerning their "prevalence" inside the variation time frame.*

Sentiment analysis otherwise called sentiment mining assumes a critical part in deciding the slants required in different Web content. Dissecting opinions is imperative for deciding. For instance, in the event that one needs to purchase another mobile phone, a Web astute purchaser will quite often first check surveys about it keeping in mind the end goal to settle on an educated purchasing choice in view of others encounters. Student sentiment analysis is as of now an exceptionally critical pattern in the zone of natural language processing. Common dialect preparing includes giving computerized reasoning to PCs and is worried with advancing a comprehension of human dialects for machines' utilization. Student sentiment analysis extricates opinions, notions, and sentiments from content and analysiss them this data is exceptionally helpful for governments, organizations and people. While this substance intended to be useful in dissecting this heft of user produced substance is troublesome and tedious. So require emerges to build up a savvy framework which mine such gigantic substance and arrange them into Negative, Positive, Neutral sort. Student sentiment analysis is the computerized mining of sentiments, states of mind, sentiments from content, and database sources through Natural Language Processing (NLP).

***Keywords:** Student sentiment Analysis, Public sentimental Social Sites, Twitter, Emerging topic mining.*

I. INTRODUCTION

With the Extensive development of user created messages on web, Social webpage like Twitter where a great many users used to impart their insight in regards to some theme. Figure 1 demonstrates that web has enormous measure of information and interpersonal organizations have part of that tremendous information. We can Sentiment analysis on Social Sites information has given a stage where opportune student Sentiment can uncover in a conservative and viable way, which is monotonous for basic leadership in different areas. For instance, an organization could investigate the general population notion in Tweets to acquire users' input towards its items or administration; while a political pioneer can conform his/her position regarding the sentiment change of people in general, sentiment about films can be most valuable for progression of motion pictures

Sentiment. Because of the gigantic of online networking services there is incredible chance to comprehend and examine the slant of the students by dissecting its huge scale information and additionally sentiment rich information. Sentiment analysis on tweets should be possible by numerous methodologies. Different strategies, for example, machine-learning and Lexicon based methodologies have been generally utilized for Sentiment analysis on Twitter like destinations. Machine-learning ways to deal with student sentiment analysis need to prepare the information.

Searching for student's opinions by means of overviews and surveys has been a costly and tedious assignment. The proliferation of Web 2.0 has changed the way students express their opinions and emotions. This supposed user created content posted in websites, gatherings, item survey locales and informal communities is for the most part publicly accessible and simple to get. Accordingly, there is a developing requirement for mechanized analysis of this sort of information. This is a testing undertaking with establishments in normal dialect preparing and message mining alluded to as slant analysis. Numerous analysis concentrates on in opinion analysis are worried with item audits from sites like Twitter is a most famous overall social site, which gives a miniaturized scale blogging services and long range informal communication, empower its users to upgrade their status in tweets, take after the general population they are keen on (e.g. Sachin Tendulkar) and retweet other's posts and even speak with them straightforwardly. The student wistful analysis on Twitter information has given a prudent and compelling approach to uncover convenient student opinions, which is basic for basic leadership in different spaces regions. Case in point, a colleges can think about the Sentiments of student in Tweets to acquire users 'input towards its training framework. There are a few floods of analysis researching the part of Twitter. Twitter has pulled in consideration in both the educated community and industry for Research Area. Past analysis fundamentally centered around following public sentiment.

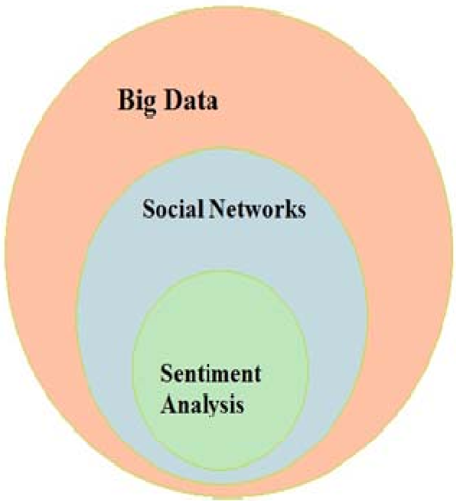


Fig. 1 Visualization of Social Network Analysis

There have been an expansive number of exploration studies and modern applications in the range of public notion following and displaying. Past exploration like O'Connor [1] concentrated on following public sentiment on Twitter and considering its connection with purchaser certainty and presidential employment endorsement surveys. On Twitter, any user can distribute a message alluded to as tweet, which is obvious on general public show.

Comparative sorts of studies have been accomplished for exploring the impression of public notions on oil value lists and securities exchanges. They reported that occasions, in actuality, sudepend have a critical and prompt impact on general public sentiment on Twitter. One significant analysis is to discover conceivable purposes for notion variation, which can give imperative data to basic leadership. E.g. in the event that antagonistic public slant towards Barack Obama increments altogether, the White House Administration Office might be willing to know why individuals have changed their sentiment and after that respond as needs be to turn around this pattern. Another illustration is, Analyzing general conclusion variation surveying for Exit survey for any Election.

II. LITERATURE SURVEY

A. Representation and Communication: Challenges in Interpreting Large Social Media Datasets (M. Rost, L. Barkhuus, H. Cramer, and B. Brown, 2013)

Online service give a scope of chances to comprehension human conduct through the vast aggregate information sets that their operation gathers. However the information sets they gather don't un-problematically model or mirror the world occasions. In this paper author utilize information from Foursquare, a well-known location check-in administration, to contend for the significance of breaking down social media as an informative as opposed to representational framework. Drawing on logs of every one of the foursquare check-in more than eight weeks author highlight four components of Foursquare's utilization: the relationship amongst participation and check-in, occasion check-in, business motivations to check-in and in conclusion funny check-in. These focuses indicate how huge information analytics is influenced by the end user uses to which social network are put.

B. The State of Educational Data Mining in 2009: A Review and Future Visions (R. Baker and K. Yacef, 2009)

These papers survey the history and current patterns in the field of Educational Data Mining (EDM). Author consider the methodological profile of exploration in the early years of EDM, contrasted with in 2008 and 2009, and talk about patterns what's more, moves in the exploration led by this group. Specifically, author talk about the expanded accentuation on forecast, the development of work utilizing existing models to make investigative revelations ("disclosure with models"), and the diminishment in the recurrence of relationship mining inside the EDM people group. Author examine two ways that analysts have endeavored to arrange the assorted qualities of exploration in educational information mining research, and audit the sorts of analysis issues that these strategies have been utilized to address. The mostcited papers in EDM somewhere around 1995 and 2005 are recorded, and their impact on the EDM people group (and past the EDM people group) is talked about.

C. Microblogging in Classroom: Classifying Students' Relevant andIrrelevant Questions in a Microblogging-Supported Classroom (S. Cetintas, L. Si, H. Aagard, K. Bowen, and M. Cordova-Sanchez, 2011)

Microblogging is a popular innovation in social networking applications that gives users a chance to distribute online short instant messages (e.g., under 200 characters) progressively by means of the web, SMS, texting customers, and so forth. Microblogging can be a compelling device in the classroom and has recently increased outstanding enthusiasm from the training group. This paper proposes a novel utilization of text order for two sorts of microblogging inquiries asked in a classroom, to be specific significant (i.e., questions that the educator needs to address in the class) and insignificant inquiries. Exact results and analysis demonstrate that utilizing personalization together with inquiry text prompts preferred classification precision over utilizing question message alone. It is likewise valuable to use the relationship between's inquiries what's more, accessible address materials and in addition the relationship between's inquiries asked in an address. Besides, observational results likewise appear that the end of stop words prompts better connection sentiment amongst inquiries and prompts better order exactness. Then again, joining students' votes on the inquiries does not enhance order exactness, despite the fact that a comparative element has been appeared to be powerful in group question noting situations for evaluating question quality.

D. An Engineering Student Retention Study (C. Moller-Wong and A. Eide, 1997)

National engineering enlistment topped in the mid 1980's and, with uncommon special case, has declined or stayed level for as long as fifteen years. Generally, engineering enlistment has concentrated on new student enrollment, however as of late considerably more consideration has been coordinated toward the issue of student maintenance. Our endeavors at Iowa State University to look at maintenance issues were isolated into two sections. Stage I of the study, which is introduced in this paper, was focused to achieve a few destinations. In the first place, author needed to plan and amass an information base that would take into account singular following of students. Once a complete profile of our under studies populace had been collected, it was conceivable to recognize precisely a scope of illustrative variables. Next, utilizing the built up information base author built up a maintenance analytics instrument that would factually recommend and recognize students who are possibly at danger of whittling down. To do this author analyzed a scope of free variables against an arrangement of ward danger classes. Factual analyticss and logistic relapse strategies were performed to give a prescient model. Stage II of this study is as of now in procedure and includes a large number of weakening variables which were not considered, measured, or incorporated into Phase I. At the point when components, for example, foundation, social incorporation, mentality, and so on can be effectively measured and coordinated into the information base the exactness of our present model is required to move forward.

E. Identity Management: Multiple Presentations of Self in Facebook (J.M. DiMicco and D.R. Millen, 2007)

As the utilization of social networking sites turns out to be progressively normal, the sorts of social relationship oversaw on these destinations are turning out to be progressively various and differing. This research tries to pick up a comprehension of the issues identified with overseeing distinctive social networks through one framework, specifically taking a gander at how users of these frameworks present themselves when they are utilizing one site to stay in touch with both their past social gatherings from school and their present social associations in the work environment. To do this, author analyzed online profile pages and met representatives at an expansive programming improvement organization who every now and again utilize the site Facebook, a site essentially utilized by undergrads and youthful graduate s transitioning into the work power. The result of this underlying contextual analysis is a system for seeing how users oversee self-presentation while keeping up social connections in heterogeneous systems

III. Proposed System

A. General Architecture

Today's all Social Networking destinations have been broadly utilized for communicating opinions or sentiment in general public area with help of web. Also, Twitter has been the purpose of fascination in a few scientists in vital territories. The fundamental two-fold commitments of this paper are: (1) to the best of our study, our exploration and our insight is the primary work that tries to investigate and translate general public sentiment variations in small scale blogging services like twitter. (2) Two novel generative models are created to take care of the reason mining issue. The two proposed models are general: they can be connected to different errands, for example, discovering theme contrasts between two arrangements of records.

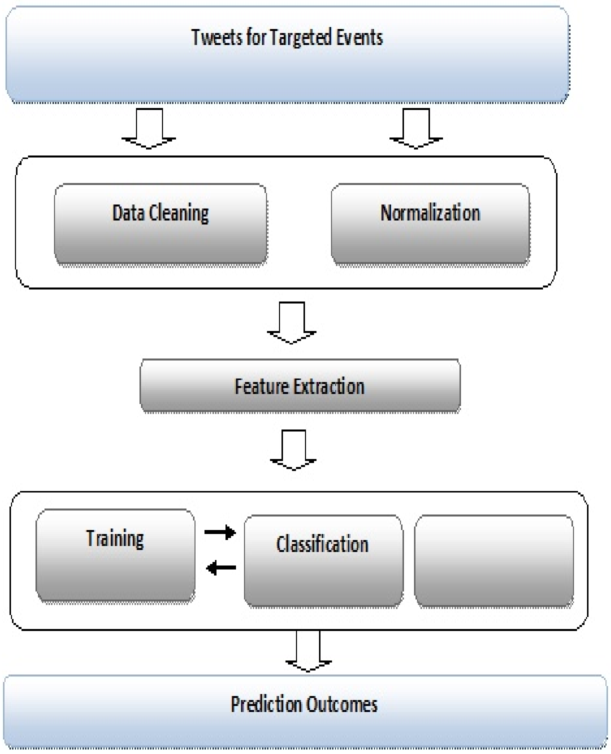


Fig. 2 High Level System Flow

Fig. 2 demonstrates a case of High level framework stream. To break down variations in broad daylight opinions. There are two Latent Dirichlet Allocation (LDA) based models: (1) Foregroundand Background LDA (FB-LDA) and (2) Reason Candidate and Background LDA (RCB-LDA). NaïveBayes, SVM, MaxEnt, ANN classifiers with components extricated from Twitter information utilizing highlight extraction strategies, for example, Unigram, Bigram and Hybrid (Unigram + Bigrams) for conclusion analysis. To expel stop words and to concentrate highlights from content, we perform information cleaning and standardization on given set. We separate the objective based developed components model [7] by changing it and twitter user information from the standardized information. Vectors are utilized as a part of piece of lumps to prepare the classifier as a piece of incremental preparing. The slant analysis results are joined with impact variable of regulated figuring out how to anticipate the outcomes utilizing forecast model.

B. Proposed Architecture

In our work, we have proposed taking after three stages for slant following.

- We separate tweets identified with our intrigued targets (e.g. Arvind Kejrival, Delhi Election 2015 and so on), and preprocess the crude separated for more Cleaned for Sentiment analysis.
- Second, we relegate some name supposed notion name for each individual tweet by consolidating two cutting edge slant analysis instruments [9], [8].
- At last, depend on the sentiment labels acquired for every tweet, we recognize the sentiment variation for the comparing focused on issues by utilizing some engaging measurements.

C. Modules

1. *Tweets Extraction and Preprocessing:* Our First stage begins with extricating tweets lines identified with the focused on issue, we experience the entire gathered crude dataset and concentrate all the center lines tweets which contain the catchphrases of the focused on issues. Contrasted and consistent content archives, tweets are for the most part fairly Informal and frequently written in an adhoc way like it might contains short structures, some shortening. Sentiment analysis can devices connected on crude tweets however regularly accomplish extremely poor execution by and large. Subsequently there is need of preprocessing systems on tweets are essential for getting agreeable results on slant analysis:
   - Slang words interpretation: The most widely recognized Tweets regularly contain a considerable measure of slang words (e.g. lol, omg). These words are normally imperative for sentiment analysis, however may not be incorporated into root notion vocabularies. Since the opinion analysis depends on sentiment Lexicon, in this manner we are changing over these all slang words into their standard structures utilizing the Internet Slang Word Dictionary and afterward re-add them to the tweets.
   - Non-English tweets sifting: Since the slant analysis instruments to be utilized work for English writings, we expel all non-English tweets ahead of time as these non-English words doesn't have importance for sentiment. A tweet could be dealt with as non-English tweet if more than 20 percent of its words (after slang words interpretation) don't show up in the GNU Aspell English Dictionary.
   - URL expulsion: A loads of users may incorporate different URLs in their tweets. These URLs may muddle our sentiment analysis process. So we choose to expel URLs from tweets. (100%)
2. *Sentiment Label Assignment:* For doling out Sentiment names for every tweet all the more unhesitatingly, we sort dictionaries again to two cutting edge notion analysis instruments. One is the SentiStrength3 instrument [8]. This device depends on the LIWC [10] sentiment Lexicon. It works in the accompanying way: first relegate an opinion

score to every word in the content as indicated by the sentiment dictionary; then pick the greatest positive score and the most extreme negative score among those of all individual words in the content; register the total of the most extreme positive wistful score and the greatest negative nostalgic score, signified as Final wistful Score; at long last, utilize the indication of Final Score to show whether a tweet is sure, negative or it is impartial.

D. Lexicon based Techniques

In unsupervised procedure, grouping is finished by looking at the elements of a given content against Sentiment dictionaries whose sentiment qualities are resolved before their utilization. Notion dictionary contains arrangements of words and expressions used to express individuals' subjective sentiments and opinions. For instance, begin with positive and negative word vocabularies, investigate the archive for which conclusion need to discover. At that point if the report has more positive word dictionaries, it is sure, else it is negative. The Lexicon based systems to Sentiment analysis is unsupervised learning since it doesn't require earlier preparing keeping in mind the end goal to group the information.

The steps of the Lexicon based procedures are underneath

i. Preprocess every crude tweet content (i.e. evacuate HTML labels, boisterous characters).

ii. Introduce the aggregate content sentiment score: s _ 0.

iii. Tokenize content. For every token, check if tokens are available in an Sentiment word reference of preparing set.

(a) If token is available in lexicon,

- In the event that token is certain, then s _ s + w.
- In the event that token is negative, then s _ s − w.

iv. Take a gander at total content sentiment score s,

(a) If s > limit, then characterize the content as positive

(b) If s < limit, then arrange the content as negative.

IV. EXPERIMENTAL RESULTS

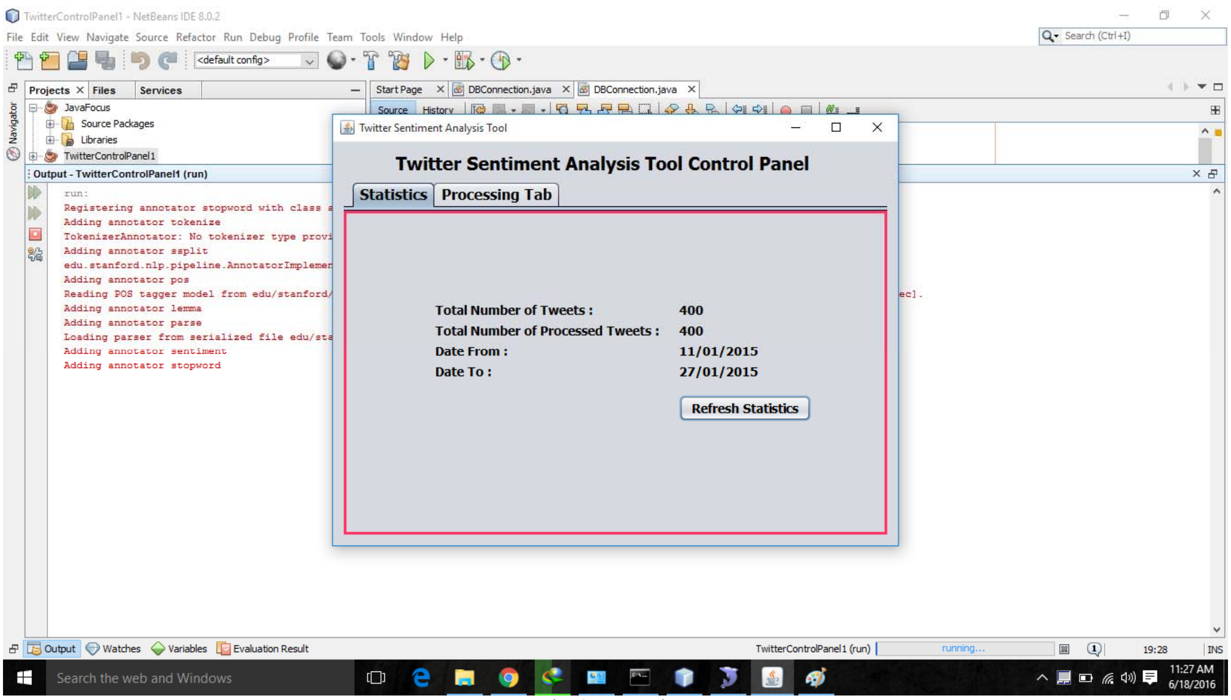


Fig. 3 Processing

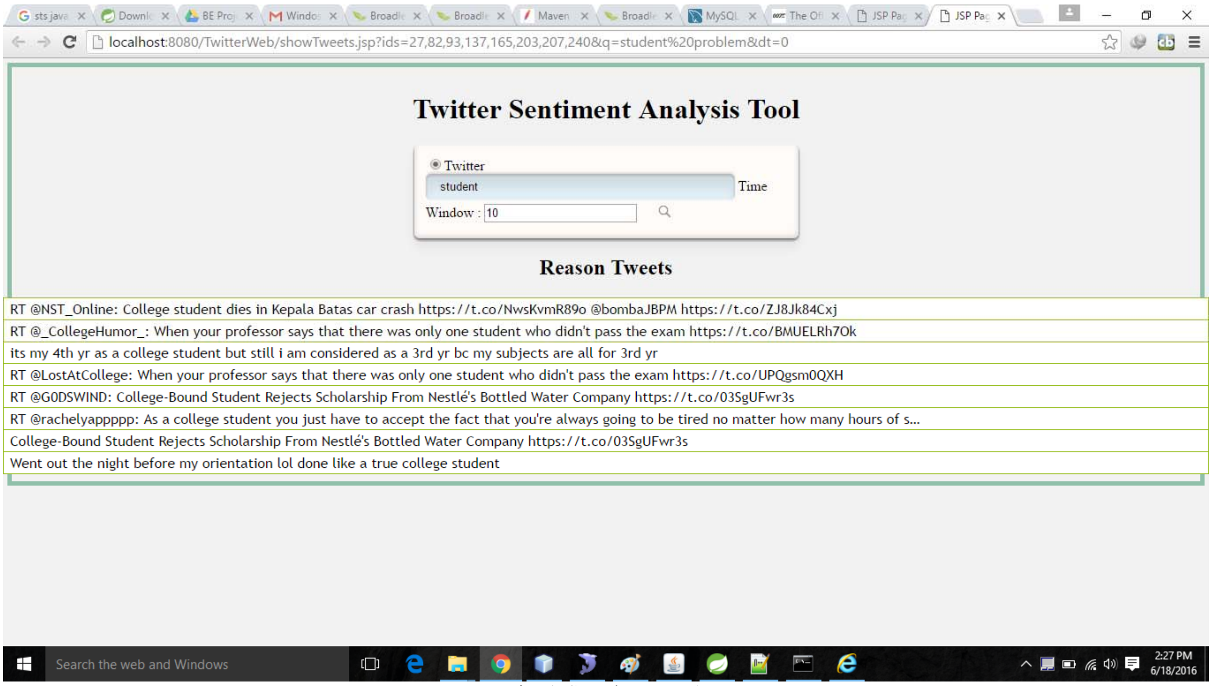


Fig. 4. Search Tweets1

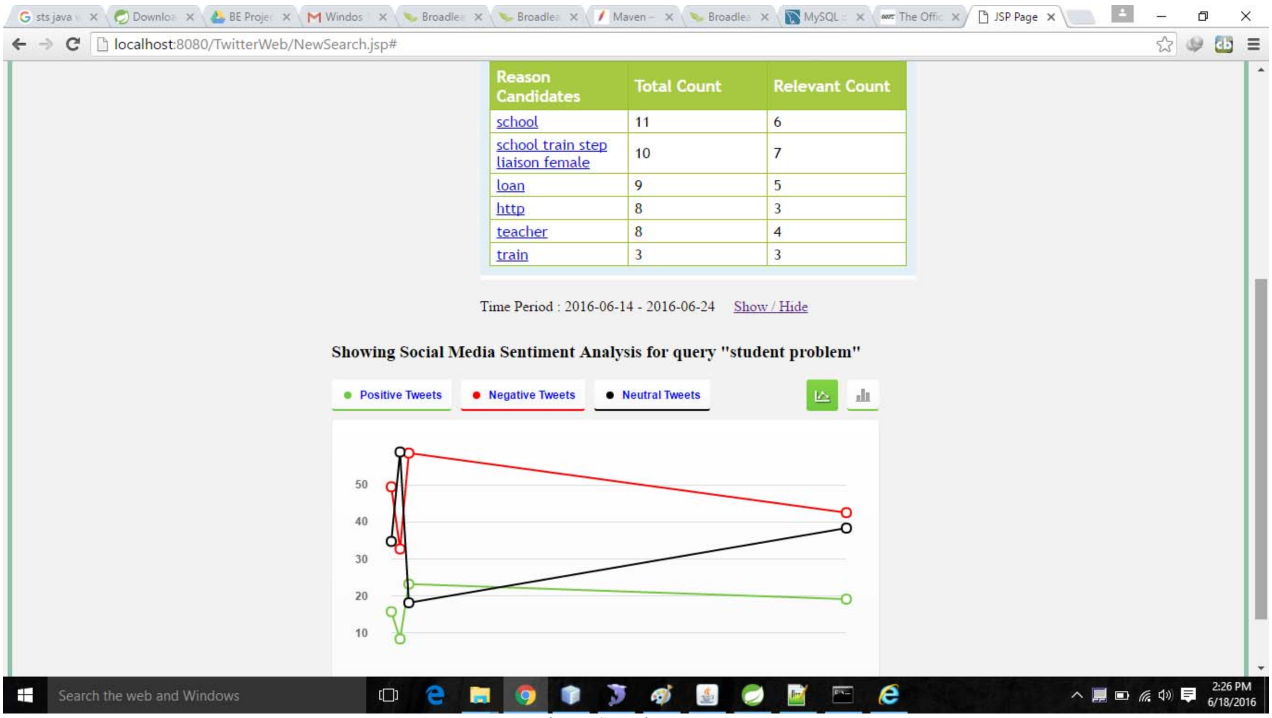


Fig. 5. Search Tweets2

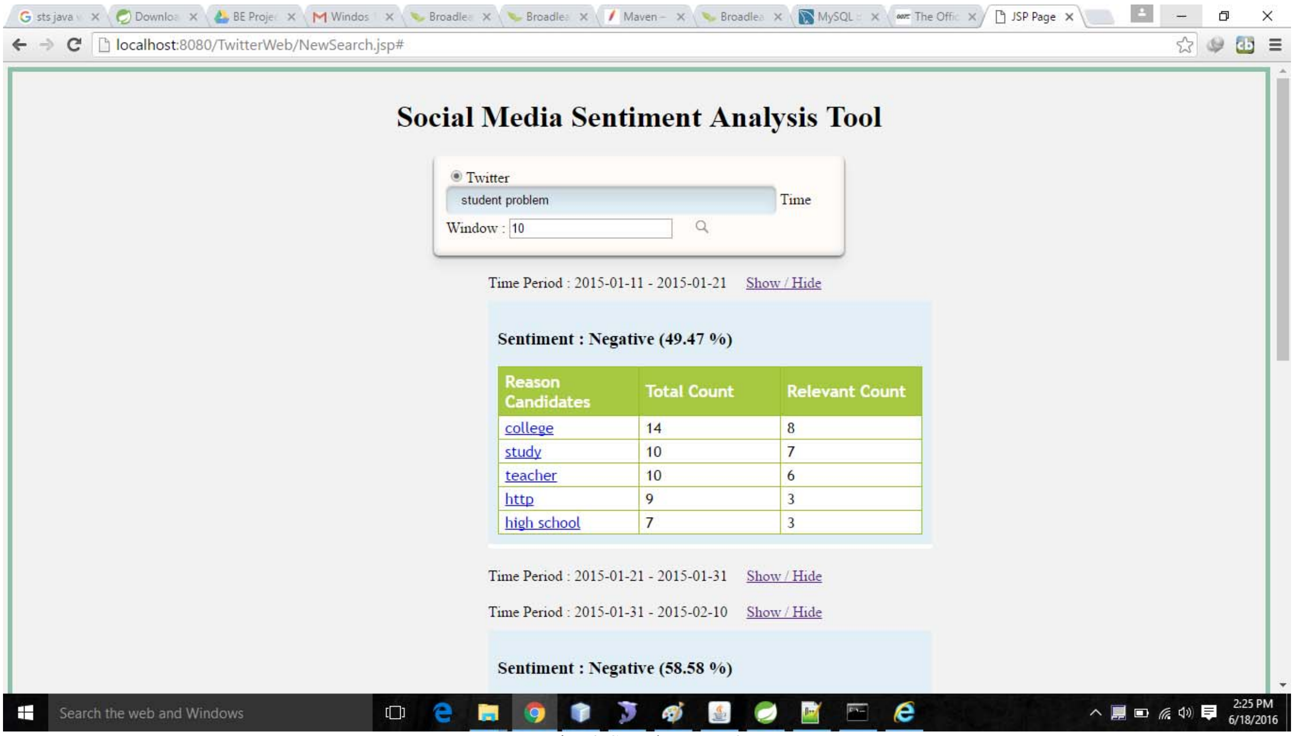


Fig. 6. Search Tweets3

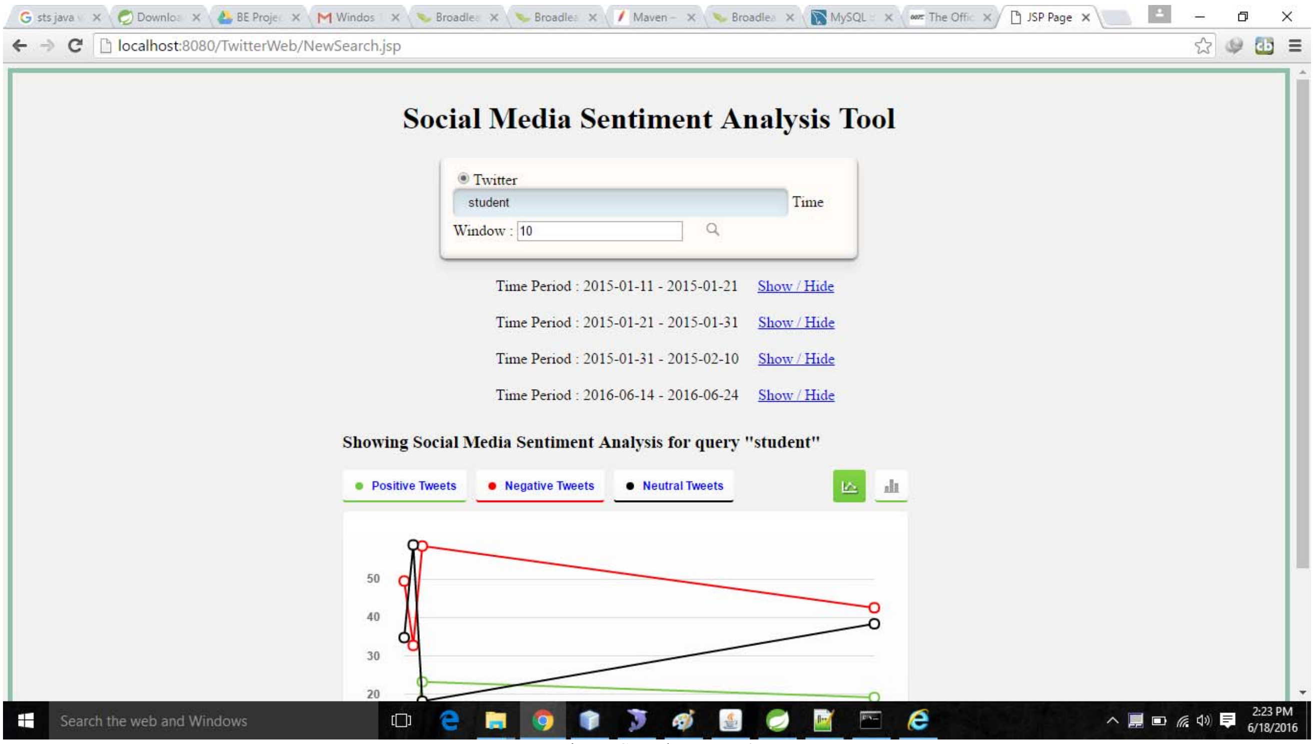


Fig. 7. Search Tweets4

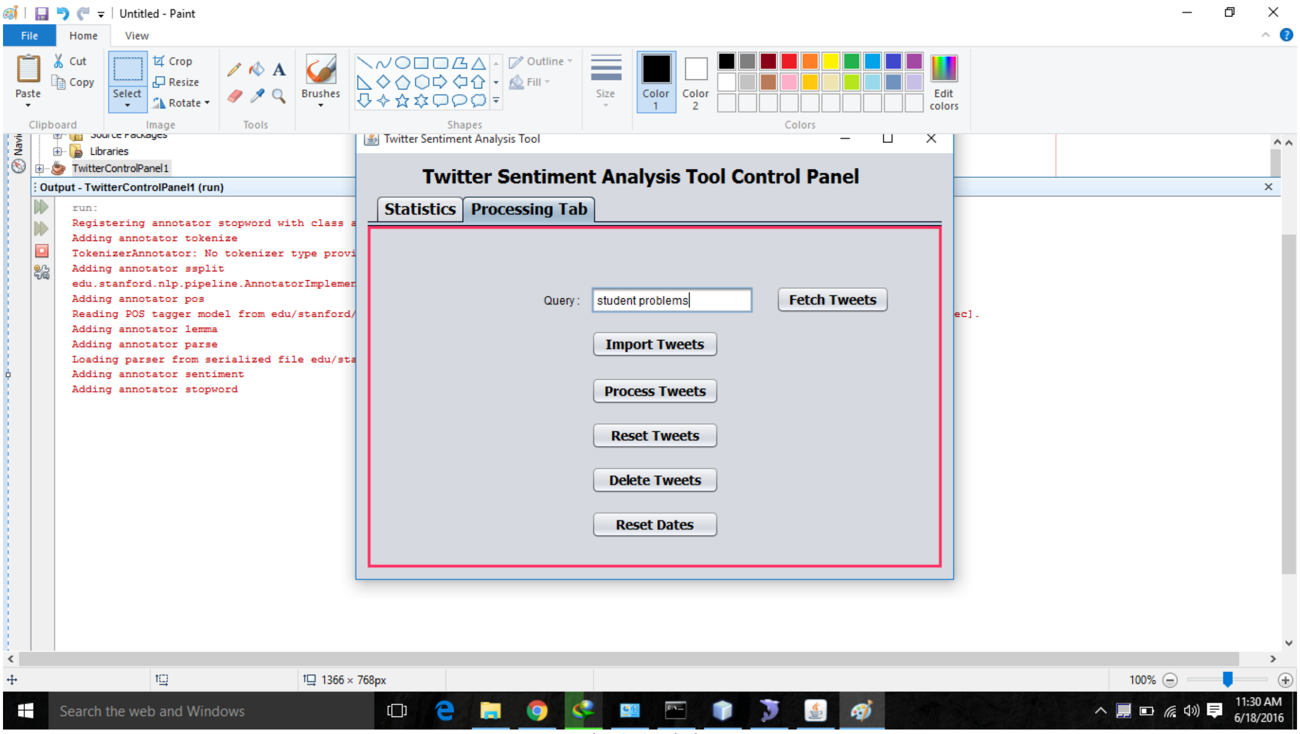


Fig. 8. Statistics

V. CONCLUSION

Generally, we presume that social network based behavioral analysis parameters can expand the prediction accuracy. Notwithstanding, presence of the considerable number of elements in impartial and equivalent way is important to give precise results. In this paper, we examined the issue of investigating student Sentiment variations and finding the conceivable reasons creating these instructive issue and overcome it. Our study can illuminate instructive chairmen and other applicable leader to increase further comprehension of building student issues. Our motivation to accomplish profound comprehension of student learning encounters.

ACKNOWLEDGEMENT

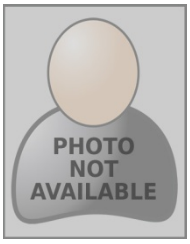
Author would like to take this opportunity to express our profound gratitude and deep regard to my ***Prof. Ajay Gupta,*** for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout my project work. His perceptive criticism kept me working to make this project in a much better way. Working under him was an extremely knowledgeable experience for me.

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.Bo Pang. Lilliam Lee, "Seeing Stars: Exploiting class relationships for sentiment categorization with respect to rating scales", 2002.
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. 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.
5. J. Bollen, H. Mao, and X. Zeng, ―Twitter mood predicts the stock market,” J. Comput. Sci., vol. 2, no. 1, pp. 1–8, Mar. 2011.
6. Bernard J. Jansen, Mimi Zhang, Kate Sobel and AbdurChowdury,Micro-blogging as online word of mouth branding”, 27th International Conference Extended Abstracts on Human Factors in Computing Systems, New York, 2009,pages 3859-3862.
7. J.C. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou.Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews”,Advances In Knowledge and organization, 2004, pages 49-54..

8. T. Minka and J. Lafferty, ―Expectation-propagation for the generative aspect model□, in Proc. 18th Conf. UAI, San Francisco, CA, USA, 2002.
9. M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, ―Sentiment strength detection in short informal text□ ,J. Amer. Soc.Inform. Sci. Technol., vol. 61, no. 12, pp. 2544–2558, 2010.
10. Nat'l Academy of Eng., The Engineer of 2020: Visions of Engineering in the New Century. National Academies Press, 2004.

AUTHOR(S) PROFILE



Sushant Patil, perceiving the M.E. degree in Computer Engineering from IOK College of Engineering, Pune, Maharashtra, India.