



“Analyzing students activities and experiences by mining social media data”

¹Supriya Sumbe, ²Pragati Lendghar, ³Priyanka Pachpute, ⁴Komal Mundra, ⁵Prof. Chhaya R.Jadhav

^{1,2,3,4,5} Department of Information Technology, D. Y. Pimpri, Pune - 412 105

Abstract --- Student informal deliberations on online network hut glow keen on top of their instructive experience. Their opinion, emotion, and concern about the knowledge process. Information on or after such uninstrumented ecological can give precious acquaintance to update student learning. Analyzing down such data, on the other hand, can be difficult. The involvedness of student experience reflects beginning social media matter require human transformation. Then again, the on the increase size of data demands usual data analysis methods. In this paper, we urbanized work procedure to integrate both qualitative psychiatry and large-scale data mining. We concentrated on engineering students' Twitter post to understand difficulty and issue in their instructive experience. We originally conducted a qualitative psychoanalysis on sample taken as of around 25,000 tweets related to engineering students' life.

Keywords --- Education, computers and education, social networking, web text analysis

I. INTRODUCTION

Information taking out investigate have successfully shaped several modus operandi, tools, and algorithms for organization enormous amount of information to respond real-world problem. As public medium is extensively second-hand for various purpose, vast amount of user shaped data be in attendance and can be completed accessible for data mining. Data mining of common medium can increase researcher's skill of thoughtful ground-breaking practice, to the exercise of common medium and increase production cleverness to in attendance high-quality services and extend original opportunity. Major objectives of the data mining course of action are to cooperatively handle significant data, take out actionable pattern, and gain understanding information. Communal medium sites such as Twitter, Face book, and YouTube present grand place to students to share happiness and struggle, sentiment and tension, and gain social support. On top of various public medium sites, student talks about their each day encounter in a happy and unceremonious manner. This Student's digital information gives huge amount of implicit information and a whole new viewpoint for educational researchers to know student's experiences outside the prohibited classroom environment.

II. LITERATURE REVIEW

Sr. No.	Paper Name	Author	Published Year	Description
1	Representation and Communication: Challenges in Interpreting Large Social Media Datasets	Mattias Rost, Louise Barkhuus, Henriette Cramer, Barry Brown.	2015	Online armed forces give a variety of opportunity for sympathetic person performance from beginning to end the great combined information set so as to their procedure collect. Thus far the statistics sets they bring together do not unproblematic ally reproduction or reflect the planet measures. In this paper we bring into participate information beginning Foursquare, a accepted site register repair, to quarrel for the significance of analysing communal medium because a outgoing somewhat than representative scheme.

2	Social Learning Analytics .	Simon Buckingham Shum ¹ and Rebecca Ferguson.	2013	We suggest so as to the plan and completion of effectual communal knowledge Analytics (SLA) in attendance noteworthy challenge and opportunity designed for both do research and venture, in three central compliments. The first is that the education scenery is unusually confused at in attendance, in no minute element due to technical drivers. Online social learning is up-and-coming as a important occurrence for a diversity of reason, which we appraisal, in order to inspire the thought of communal knowledge. The subsequent confront is to recognize dissimilar type of SLA and their linked technology and use
3	In-packet Bloom filters: Design and networking applications	Christian E. Rothenberg, Carlos A. B. M., Maurício F. Magalhaesa, Fábio L. V., A. Wiesmaierc	2011	This paper explores an exciting front in the Bloom filter research space, namely the special category of small Bloom filters carried in packet headers. Using iBFs is a promising approach for networking application designers choosing to move application state to the packets themselves. At the expense of some false positives, fixed-size iBFs are amenable to hardware and present a way for new networking applications.
4	Penetrating the Fog: Analytics in Learning and Education	Siemens, George; Long, Phil	2011	Efforts toward envisage the upcoming of teaching frequently highpoint novel technologies-ubiquitous calculating plans, supple schoolroom projects, and groundbreaking graphic shows. Nonetheless the greatest affected issue determining the upcoming of advanced teaching is rather that persons cannot really trace or understand: "big data and analytics." Knowledge analytics remains motionless cutting-edge the initial phases of application and research. Many queries are everywhere how analytics narrates to current structural schemes.
5	Enabling Engineering Student Success: The Final Report for the Center for the Advancement of Engineering Education.	Atman, Cynthia J. Sheppard, Sheri D. Turns, Jennifer; Adams, Robin S. Fleming,	2012	Today's engineering graduates will solve tomorrow's problems in a world that is advancing faster and facing more critical challenges than ever before. This situation creates significant demand for engineering education to evolve in order to effectively prepare a diverse community of engineers for these challenges. Such concerns have led to the publication of visionary reports that help orient the work of those committed to the success of engineering education. Research in engineering education is central to "all" of these visions. Research on the student experience is fundamental to informing the evolution of engineering education. A comprehensive sympathetic of the manufacturing scholar knowledge includes rational around varied theoretical paths, steering of these paths, and choice points in what way scholars select manufacturing databases, circumnavigate finished their packages, and formerly transfer happening towards works and jobs.

6	The State of Educational Data Mining in 2009: A Review and Future Visions	Ryan Baker, Yacef S.J.d. Kalina	2013	We appraisal the history and present tendencies in the field of Educational Data Mining (EDM). We reflect the procedural profile of investigation cutting-edge the early centuries of EDM, associated to in 2008 and 2009, and deliberate tendencies and changes in the investigation lead by this public. In specific, we deliberate the augmented stress on forecast, the appearance of effort by current replicas to brand technical detections (discovery with models), and the discount in the incidence of association removal inside the EDM public. We deliberate binary habits that investigators.
---	---	---------------------------------	------	---

III. PROPOSED SYSTEM

In the proposed system the researchers contain be second-hand the usually method such as survey, interview, center group, and classroom behavior to bring together information associated to students' education experience. These method be frequently incredibly protracted, consequently cannot be duplicate or repetitive by way of high incidence. Frequently the examination of such study is also inadequate. In adding, when encouraged in family member to their experience, student needs to return on could you repeat. They were accepted wisdom and responsibility a moment or two in the past, which may encompass become covered over point in time.

3.1 Advantages of Proposed System:

1. It proposes a workflow to bridge and integrate a qualitative research methodology and large scale data mining techniques.
2. It provides deep insights into engineering student's educational experiences as reacted in informal, uncontrolled environments

IV. SYSTEM ARCHITETURE

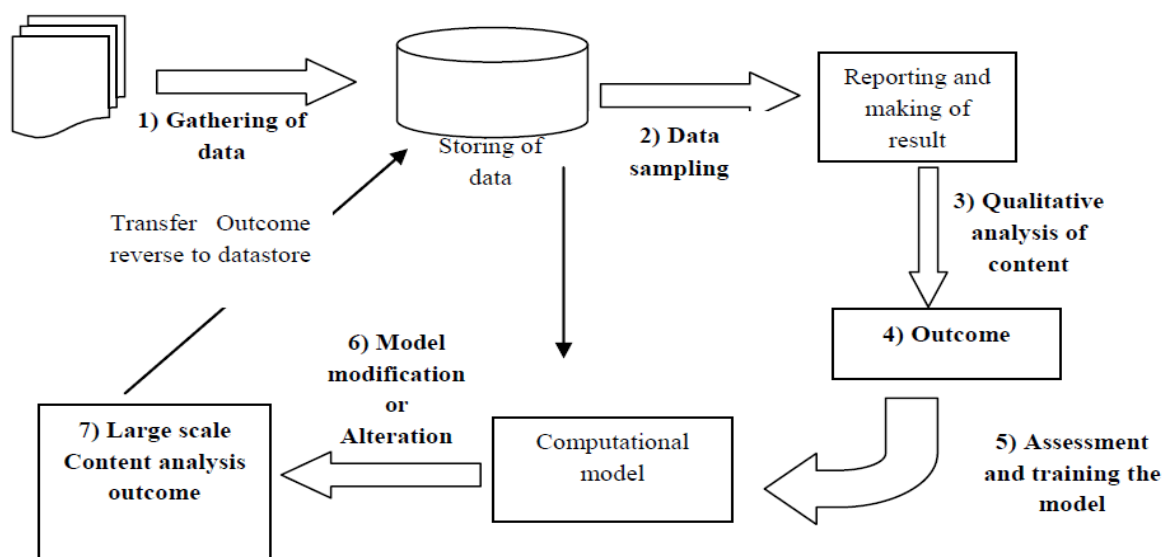


Figure 4.1 System Architecture

V. MATHEMATICAL MODEL

Let X is the Whole System Consist of

$$X = \{IP, PRO, OP\}$$

Where,

A) IP is the input given to the system

$$IP = \{N, S, W, L, C, D, R\}.$$

Where,

1. Let N be set of total number of words in a collected document i.e. tweets.

2. Let W be the set of tweets collected.

$$W = \{w_1, w_2 \dots w_n\}.$$

3. Let D be the set of number of tweets or documents collected.

$$D = \{d_1, d_2 \dots d_n\}.$$

4. Let C be the set of total number of categories of tweets.

$$C = \{c_1, c_2 \dots c_n\}$$

5. Let R be the set of researchers who are categorizing the tweets of users.

$$R = \{R_1, R_2 \dots R_n\}.$$

6. Let S be the number of labels i.e. tweets that are common between researchers.
 $S = \{S_1, S_2 \dots S_n\}.$

B) PRO is the procedure of system:

If a word w_n appears in a category C for

times, and appear in categories other than C for

times, then based on the maximum likelihood estimation, the probability of this word in a specific category C is,

$$p(w_n | c) = \frac{m_{w_n c}}{\sum_{n=1}^N m_{w_n c}}.$$

Similarly, the probability of this word in categories other than c is

$$p(w_n | c') = \frac{m_{w_n c'}}{\sum_{n=1}^N m_{w_n c'}}.$$

Suppose there are a total number of M documents (tweets) in the training set i.e. set of collected tweets, and C of them are in category c. Then the probability of category c is,

$$p(c) = \frac{C}{M}$$

and the probability of other categories c' is

$$p(c') = \frac{M - C}{M}$$

For a document d_i in the testing set, there are K words

$$W_{d_i} = \{w_{i1}, w_{i2}, \dots, w_{iK}\}$$

Where, W_{d_i} is the subset of W .

According to Bayes' Theorem, the probability that d_i belongs to category c is,

$$p(c | d_i) = \frac{p(d_i | c) \cdot p(c)}{p(d_i)} \propto \prod_{k=1}^K p(w_{ik} | c) \cdot p(c)$$

and the probability that d_i belongs to categories other than c is,

$$p(c' | d_i) = \frac{p(d_i | c') \cdot p(c')}{p(d_i)} \propto \prod_{k=1}^K p(w_{ik} | c') \cdot p(c')$$

Hence,

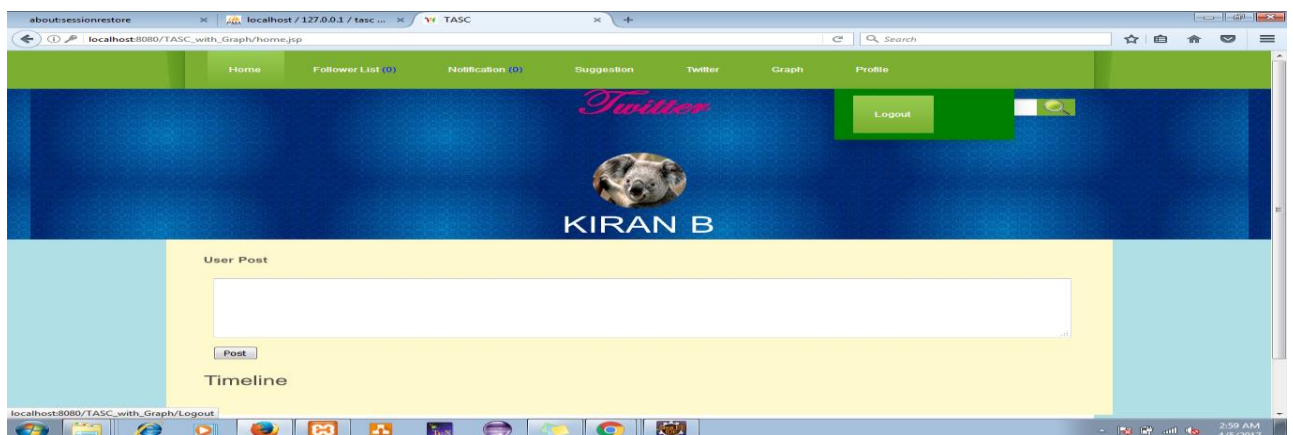
c) OP is the Output of the system:

$$P(c|d_i) + P(c'|d_i) = 1.$$

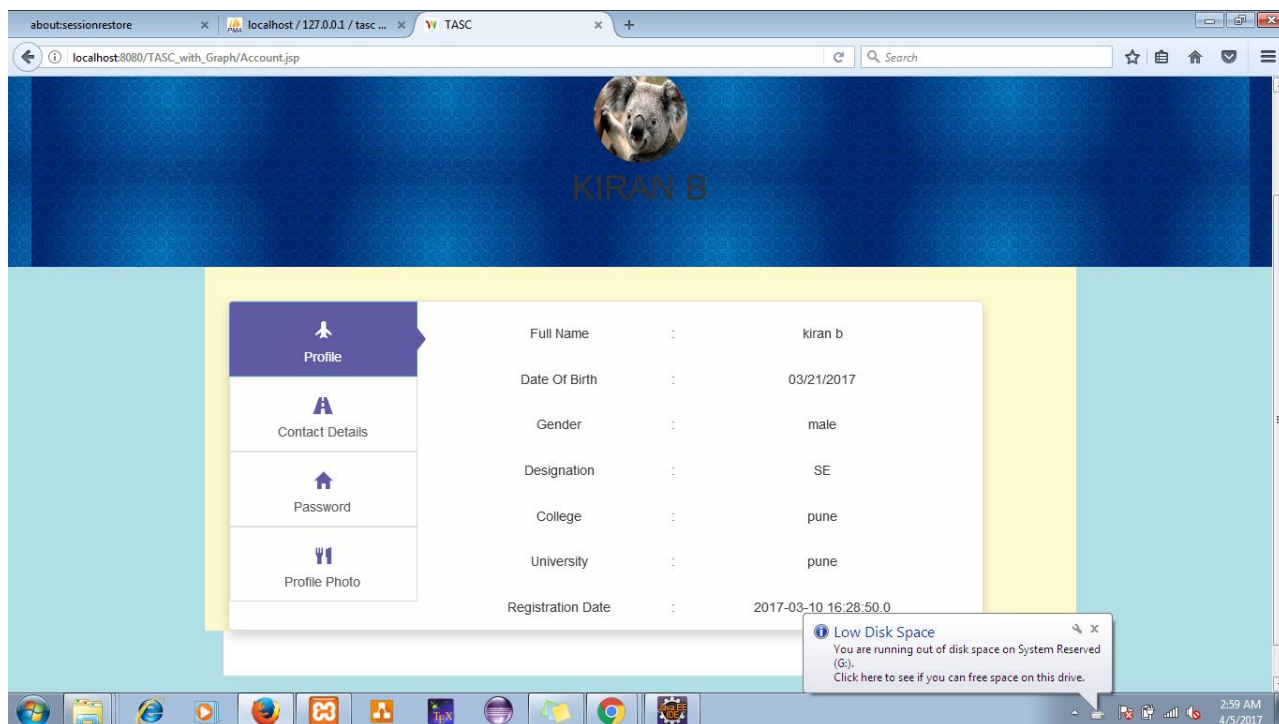
We normalize the latter two items which are proportional to $P(c|d_i)$ and $P(c'|d_i)$ to get the real values of $P(c|d_i)$ which is our objective.

VI. RESULT ANALYSIS

Screenshot 1:



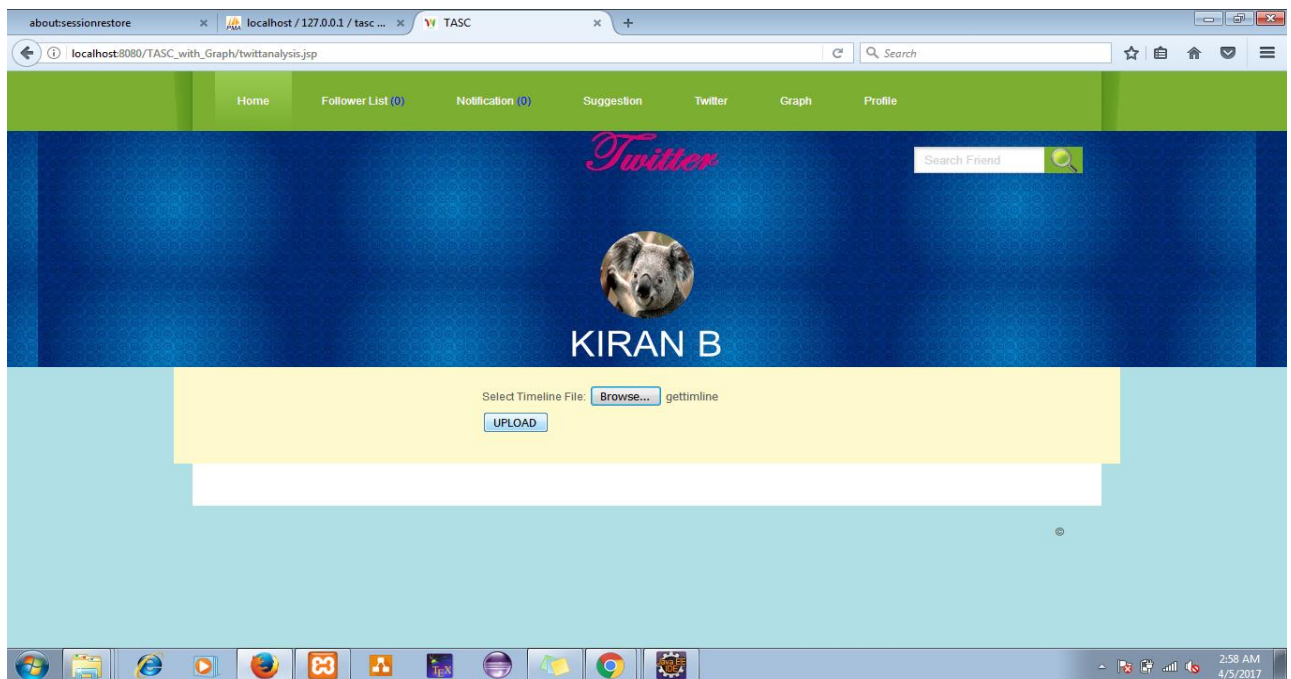
Screenshot 2:



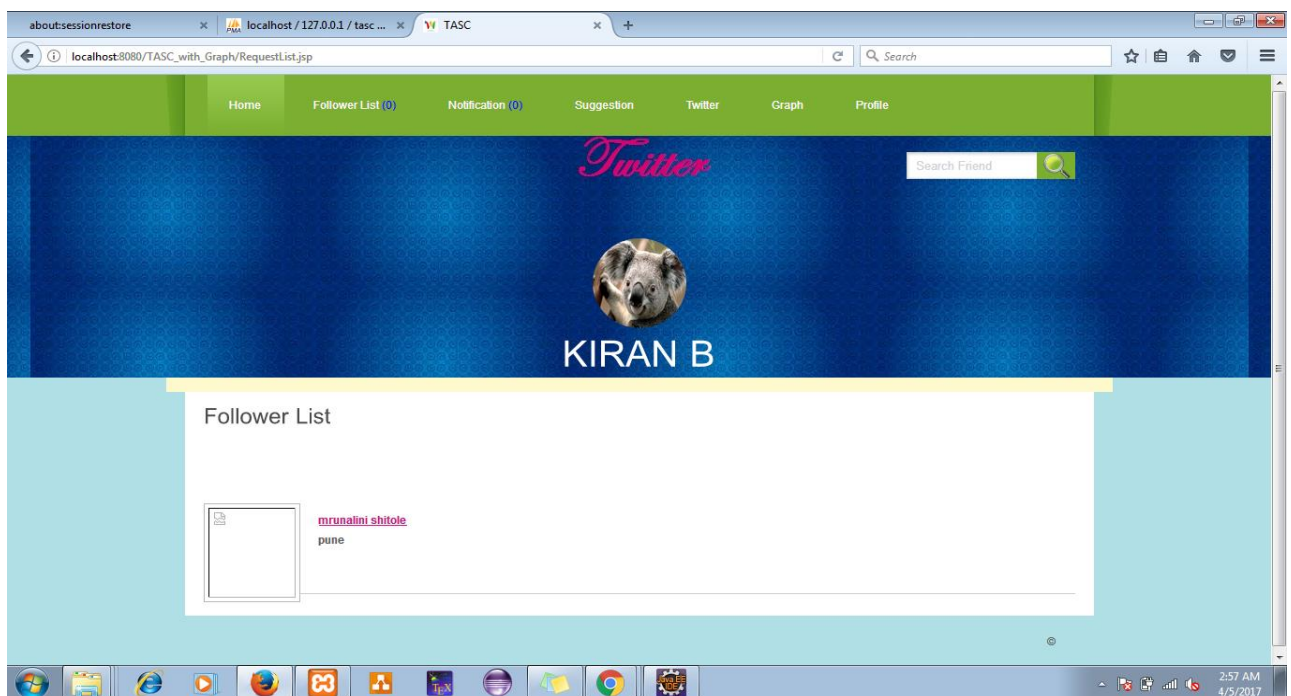
Screenshot 3:



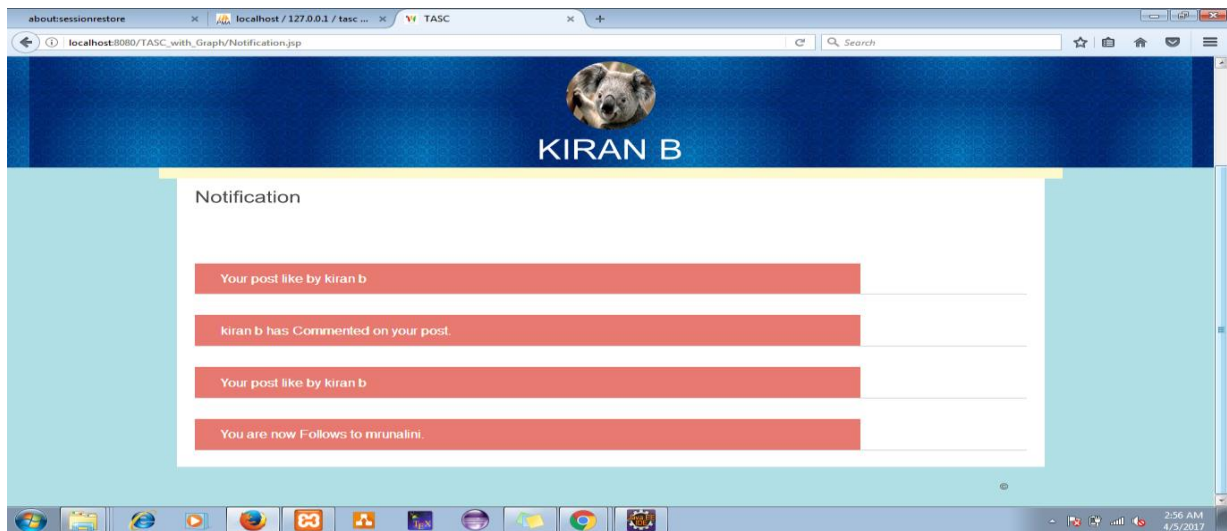
Screenshot 4:



Screenshot 5:



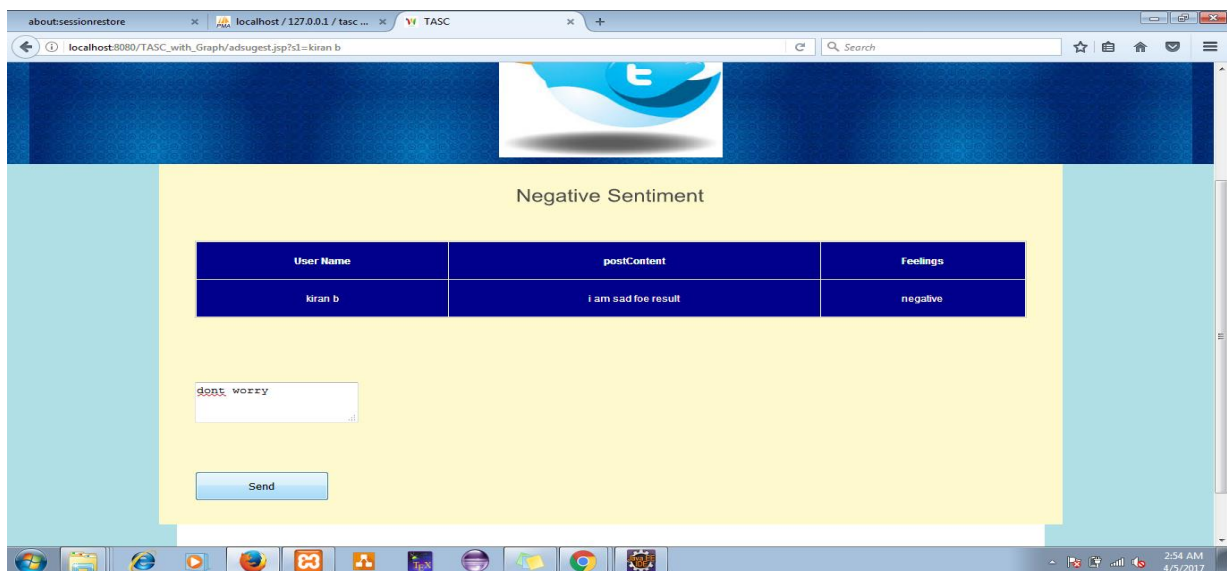
Screenshot 6:



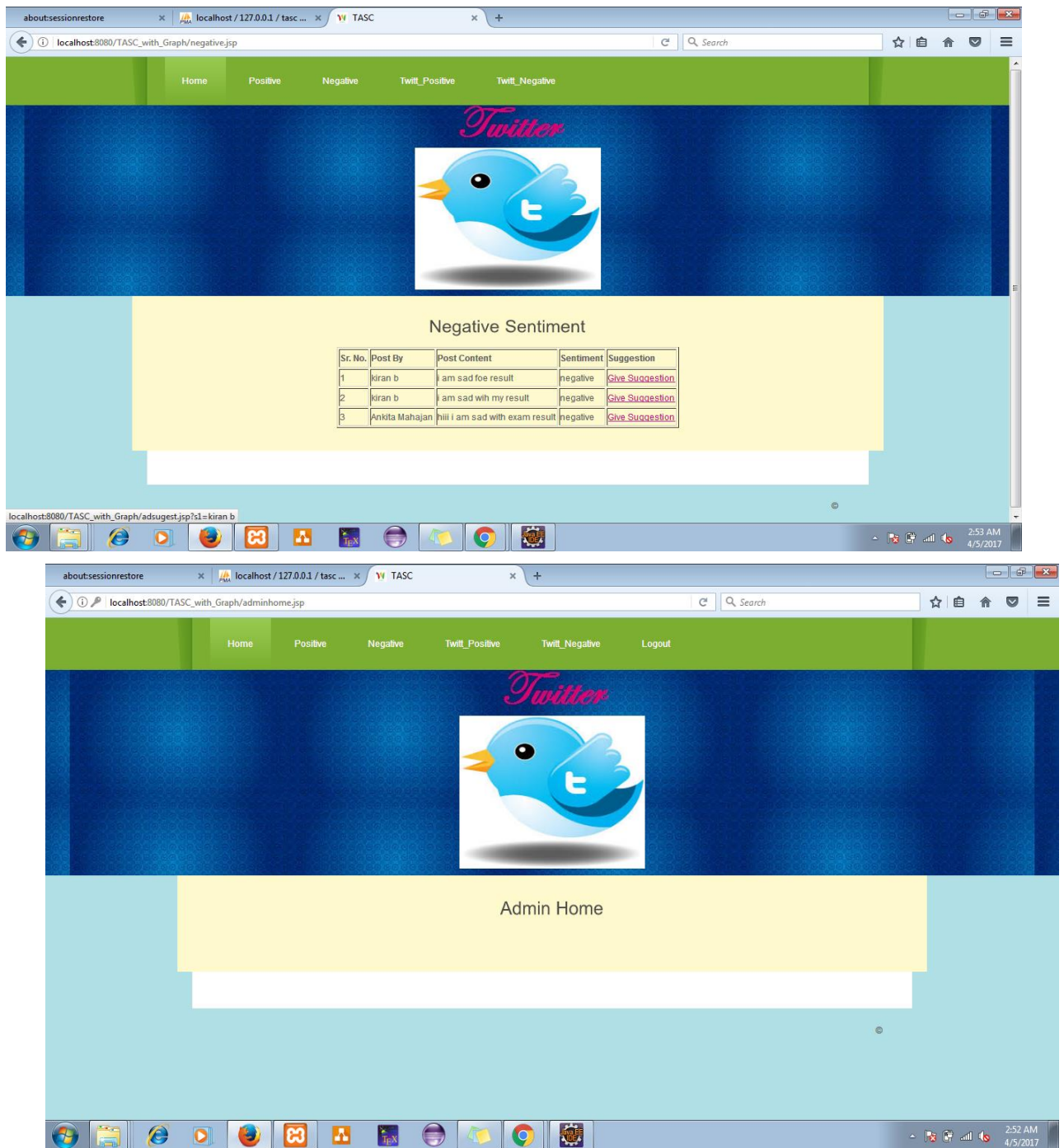
Screenshot 7:



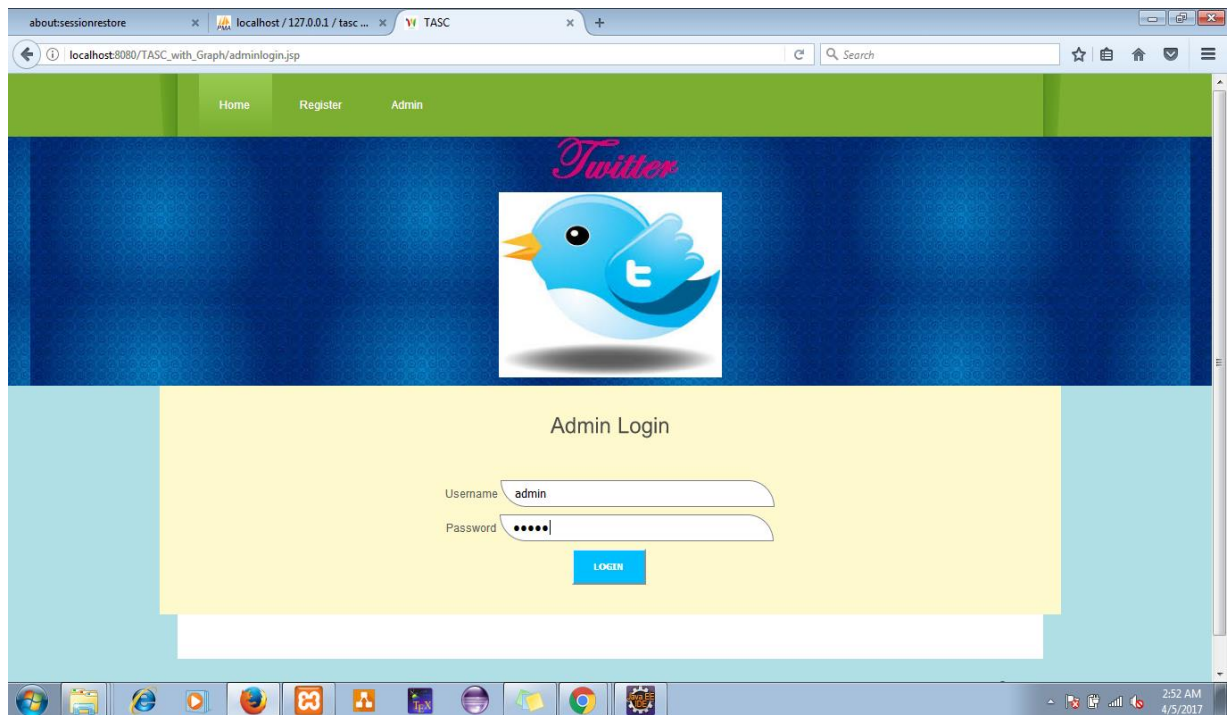
Screenshot 8:



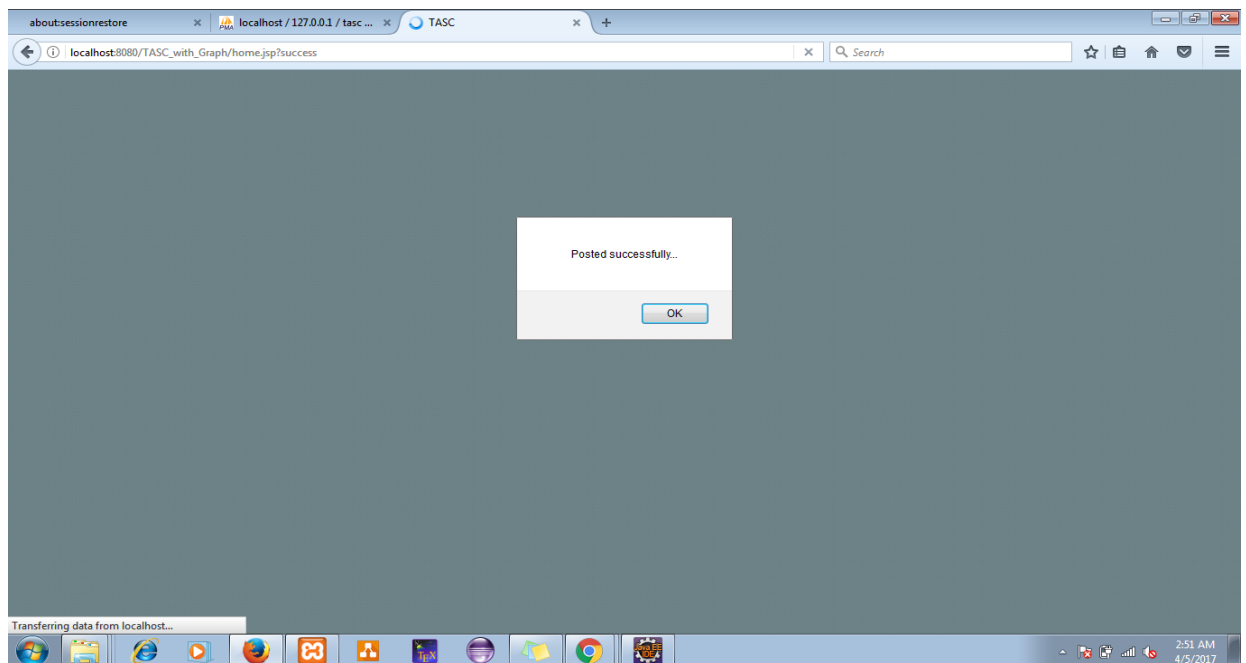
Screenshot 9:



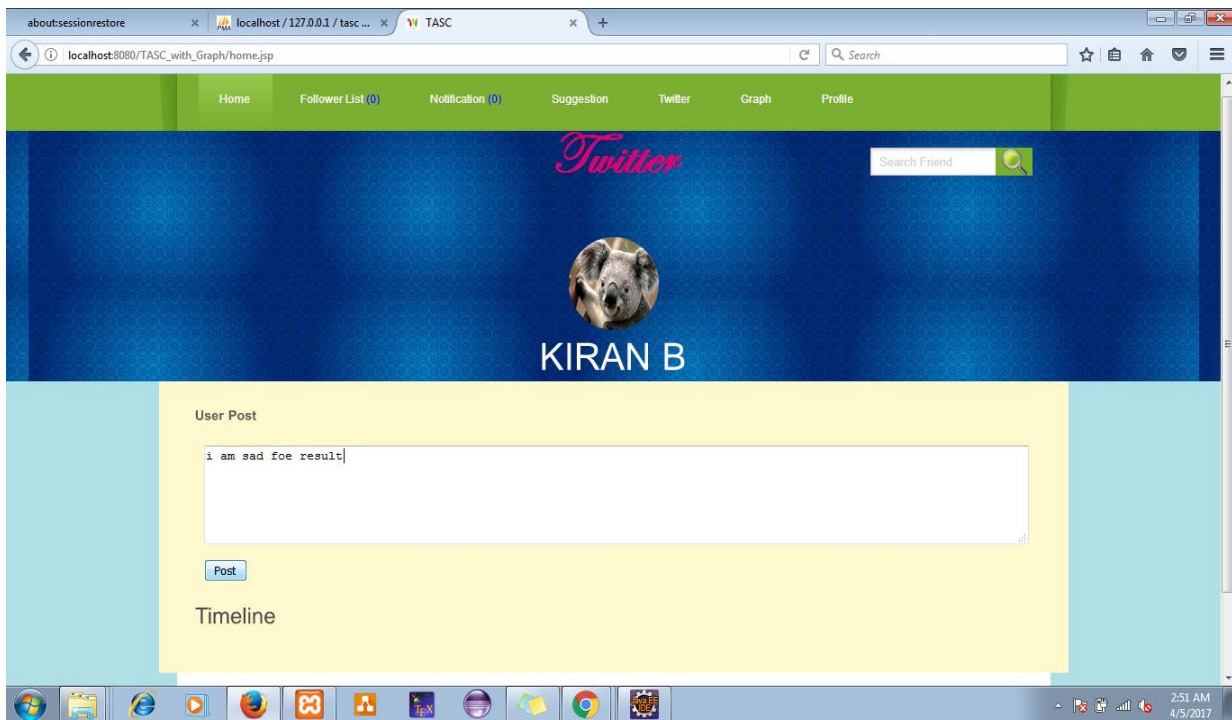
Screenshot 10:



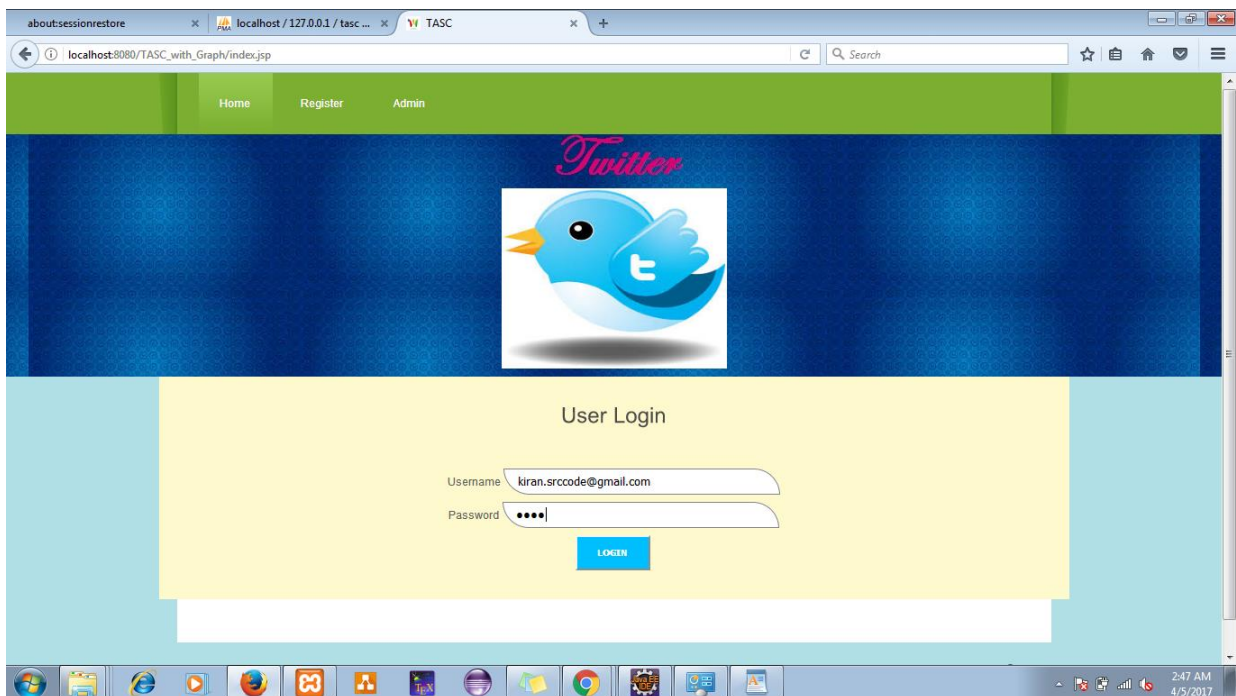
Screenshot 11:



Screenshot 12:



Screenshot 13:



VII. CONCLUSION AND FUTURE SCOPE

Mining social media information is obliging to researchers inside knowledge analytics, instructive information elimination, and knowledge ability. It give a method to investigative communal average figures that overcome the most important restrictions of both physical qualitative examination and enormous scale computational learning of user shaped textual satisfied. Two algorithms are helpful intended for this categorization primary the inexperienced Bayes Multi-label Classifier and next the hierarchy building algorithm. It notifies instructive director, and additional appropriate appraisal makers to get bigger additional compliant of manufacturing students' organization sympathetic.

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.
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. R. Ferguson, "The State of Learning Analytics in 2012: A Review and Future Challenges," *Technical Report KMI-2012-01*, Knowledge Media Inst. 2012.