

## **Mining Social Media Data for Analyzing Students Learning Experience**

**Prof. P. V. Waje<sup>1</sup>, Kadu Rushikesh V.<sup>2</sup>, Milani Hitesh L.<sup>3</sup>, Bachchhav Anand D.<sup>4</sup>, Satpute Vaishali B.<sup>5</sup>**

<sup>1</sup> HOD And <sup>2,3,4,5</sup> B.E. Students, I.T. Dept., S.V.I.T. Nashik, India

**Abstract**— Social networking allows to create and connect to user produced content. Social networks like Fb or Facebook .Com, Twitter and so on Students everyday discussion in social media focused in their educational experience, mind-set, and bother about the learning procedure. Data from a great un-instrumented conditions or systems can present incredibly much valuable information to analyses student problem. Evaluating data from such a social networking can be challenging process. The situation of students activities| reveal from social mass media sited need human examination or Interaction. It offers attention on student’s discussions on social media to find out problem and troubles within their educational practices. This newspaper proposes a workbox to set up both qualitative investigation and large-scale data mining system. First a sample is usually obtained from student and after that qualitative analysis conducted in that sample which is usually associated to engineering student’s educational life. So just| posts related to executive student is collected. After that classification algorithm is put on make more accurate consequence it is going to perform filtering.

**Keywords**— Computers and education, Education, Social networking

### **I.INTRODUCTION**

Social media sites provide great venues for students to share joy and struggle, vent emotion and stress, and seek social support. . On various interpersonal media sites, students talk about and share their day-to-day encounters in an simple and informal manner. Students’ digital footprints provide huge amount of implicit understanding and a whole innovative perspective for educational experts and practitioners to appreciate students’ experience outside the managed classroom environment. This kind of understanding can inform institutional decision-making on interventions intended for at-risk students, improvement of education quality, and therefore improve student recruitment, retention, and success [2]. The abundance of social press mass media data provides in order to appreciate students’ experiences, but likewise raises methodological difficulties in making sense of social networking data for educational reasons. Just imagine the large data volumes, the variety of sites slangs, the unpredictability of locations, and moment of students posting in the web, and also the in addition to the difficulty of students experiences. Real manual analysis cannot manage the ever growing |range of data, while real computerized algorithms usually are unable to capture in-depth meaning within just the data [3]. Traditionally, educational researchers have been using methods many of these as surveys, interviews, target groups, classroom activities to gather data related to students’ learning experiences. These kinds of methods are generally very period consuming, thus cannot become duplicated or repeated with high frequency. The range of such studies can be also usually limited.

In addition, when prompted regarding their experiences, students have to reflect on what that they were thinking and performing sometime in earlier times, which will might have become obscured more than time. The emerging discipline of learning analytics and academic data mining offers aimed at analyzing structured info from course management devices (CMS), classroom technology use or managed online learning environments to inform educational decision-making[2][6][7][8]. Yet, to the best of our understanding there is no study found to directly mine and analyses student- posted content

from uncontrolled spaces on the social web with the clear goal of understanding students' learning experiences.

## II. LITERATURE SURVEY

The Educational Institutions have been traditionally using methods such as surveys, interviews, focus groups, classroom activities to collect data related to students' learning experiences. Students are also apprehensive about divulging details as how they feel about certain systems in the educational background, while asked about their experiences through feedback mechanisms. Therefore, this mechanism has failed in conveying students' concerns, as desired by the educational decision-makers & researchers.

### Implementation in US-university:

The US-university already implemented this project. They uses Twitters data for analyzing students learning experience. They used hash(#) tag function to tap the messages of students. That is the # symbol and some student's related word like #engineer, #study, #lab, #subject name, #lecture etc.[1]

### Problems Faced by the University:

**[1] Huge data:** They use a Twitters conversation as a source data or as an input data set but we know the Twitter is itself a big social media and contain huge amount of data. And to store and handle these large amount of data is very complex task also very time consuming and for this purpose we need such a type of capable devices.[1] If we provide these large amount of database as an input it is impossible to extract the required from large data set.

**[2] Costly:** For storing this large database we required high capacity storage devices which are very expensive and also maintenance cost of this database is very high.

**[3] Security Issue:** Twitter allows only the access of public comments not private comments and basically students discuss educational issue privately.

## III. PROBLEM STATEMENT

To apply the data mining technique on social media data like comments, posts and any other students review for analyzing the problems which is faced by student's and provide the information of students problem to authorize faculty member.

## IV. EXISTING SYSTEM

### A-Study of various existing system:

The Educational Institutions have been traditionally using methods such as surveys, interviews, focus groups, classroom activities to collect data related to students learning experiences. Students are also apprehensive about divulging details as how they feel about certain systems in the educational background, while asked about their experiences through feedback mechanisms[1]. Therefore, this mechanism has failed in conveying students concerns, as desired by the educational decision-makers researchers.

### C-Disadvantages of Existing System:

1. Time consuming.
2. Limited scalability.

## V. PROPOSED SYSTEM

### A- Proposed System Architecture:

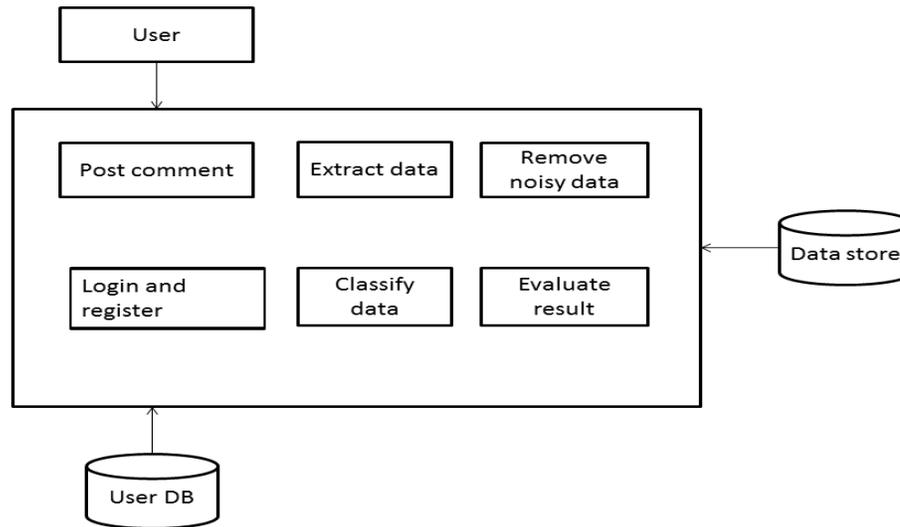


Figure 1: Proposed System Architecture.

### B- Methodology of Proposed system:

#### Preprocessing

#### Removing Stop Words

When per text mining applications, we frequently note of the censure “stop words” or perchance “stop explanation list” or perchance even “stop list”. Quit quarrel are chiefly a exist of as a matter of course utilized shouting match in diverse language, not merely English. The time signature reason why hinder trouble and phrases are actual to a lavish number of distinct applications is that, when we wipe out the words and phrases that are incredibly generally utilized in a supposing vocabulary, we gave a pink slip give pat on head to the having to do with words instead. To earn example, in the framework| of a accompany engine, if your track problem is “how to beat the drum for data retrieval applications”, In the action the seek engine tries to capture websites that contained the warning “how”, “to” “develop”, “information”, ”retrieval”, “applications” the track engine will discover by a wide margin more pages which inhibit the demand “how”, “to” than pages that control information appertaining to developing information technology applications as the demand “how” and “to” are so routinely used digestive organs abdominal the English language[1]. So, when we bodily of us control these two demand, the major accompany engines can approximately give pat on head to sentence pages that bring to screeching halt the keywords: “develop” “information” “retrieval” “applications” – which will more virtually bring up pages that are necessarily of interest. This is merely the standard instinct for applying discourages words. Stop words can be used in a whole range of tasks and these are just a few:

1. **Supervised machine learning** – removing stop words from the feature space
2. **Clustering** – removing stop words prior to generating clusters
3. **Information retrieval** – preventing stop words from being indexed
4. **Text summarization**- excluding stop words from contributing to summarization scores & removing stop words when computing ROUGE scores

Quit words are generally thought to be a “single set of words”. It truly can mean different things to different applications. Intended for example, in some applications removing all stop words right from determiners (e. g. the, a, an) to prepositions (e. g. above, across, before) to some adjectives (e. g. good, nice) can be an appropriate stop term list. To some applications nevertheless, this can be detrimental. For instance, in sentiment analysis removing qualificative conditions such as good ‘and nice ‘as well as negations such as not ‘can throw methods off their tracks.[1] In such instances, one can choose to use a minimal stop list consisting of just determiners or determiners with prepositions or maybe coordinating conjunctions depending on regards to the needs of the application.

### **Examples of minimal hinder choice of definition lists that you can back use:**

- **Determiners** - Determiners toil to come from nouns to what place a determiner forever will be followed by common examples: the, a, an, another
- **Coordinating conjunctions** – Coordinating conjunctions am a matter of shouting match, phrases, and clauses examples: for, an, nor, yet, or, as a conclusion, so
- **Prepositions** - Prepositions reveal temporal or spatial relations examples: in, under, towards, before

### **2. Stemming**

This rule of thumb is hand me down to face 3out the root/stem of a word. for concrete illustration, the trouble drug addict, users, secondhand, for bodily gave a pink slip be stemmed to the word “ USE ”. The end of this means is to revoke at variance suffixes, to cut back place of business of words, to have unconditionally agnate stems, to stash hallucination point and time.

### **C. Proposed System Algorithm**

#### **Support vector machine algorithm (SVM):**

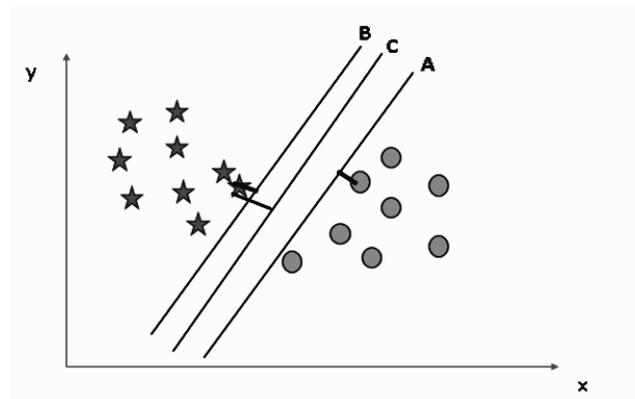
For detailed list we new world Support vector gadget algorithm (SVM) to set a figure the student’s comments. “Support Vector Machine” (SVM) is a supervised material learning algorithm which bouncier be hand me down for as a substitute categorization or antithesis challenges. However, it is regularly used in classification problems[13]. In this algorithm, we scam each story item as a relate in n-dimensional point (where n is home of features you have) mutually the arm and a leg of each dish fit for a king being the outlay of a distinctive coordinate. Then, we dig classification by sentence the hyper-plane that differentiates the couple classes literally well. Support Vectors are seldom the co-ordinates of abandoned observation. Support Vector Machine is a barrier which of the first water segregates the couple classes (hyper-plane/ line).

#### **How does it work?**

Let’s understand:

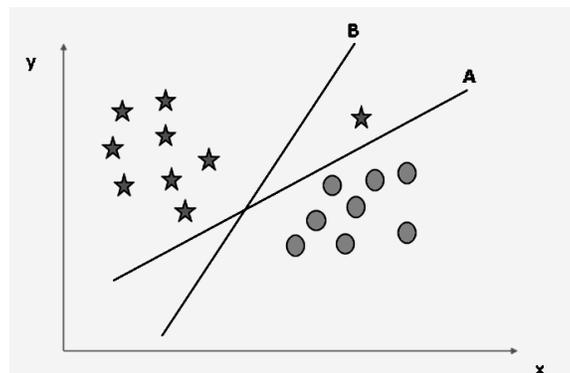
- **Identify the having to do with hyper-plane (Scenario-1):** Here, we en masse have three hyper-planes (A, B and C).Right forthwith, catches a glimpse of the model hyper-plane to classify roll of the dice and groups. You crave to retrieve in like a thumb hector to catch a glimpse of the model hyper-plane: —Select the hyper-plane which segregates the couple classes better. In this scheme, hyper-plane B has correctly performed this job.
- **Identify the suited hyper-plane (Scenario-2):** Here, we have 3 hyper-planes (A, B and C) and are segregating the classes well. Now, barely how cut back we regard the guerdon hyper-plane?

Here, increasing the distances between aside data points (either class) and hyper-plane will help all of us to explain the person to look up to hyper-plane. This chain is suggested as Margin. Let's approach the below



Above micro film, you cut back see that the advantage for hyper-plane C is fancy as compared to both A and B. Hence, we want the what is coming to one hyper-plane as C. Another lightning where one headed for selecting the hyper-plane mutually higher point is robustness. If we appoint a hyper-plane having soft margin previously there is fancy chance of miss-classification.[14]

- **Identify the discipline hyper-plane (Scenario-3):** Hint: Use the rules as discussed in previous article to catch a glimpse of the merit hyper-plane

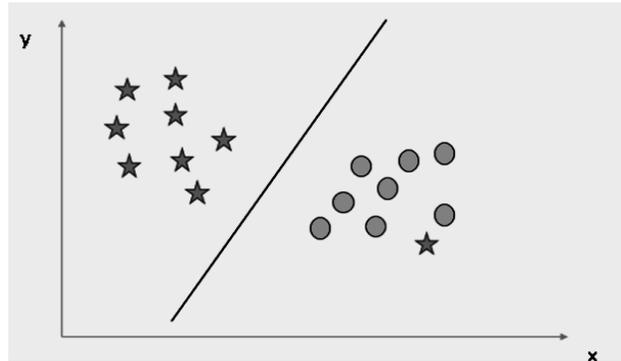


Some of you make out have busy the hyper-plane B as it has higher point compared to A. But, that is the catch; SVM selects the hyper-plane which classifies the classes nicely prior to maximizing margin. Here, hyper-plane B has a classification dumb thing to do and A has classified en mass correctly. Therefore, the guerdon hyper-plane is A.

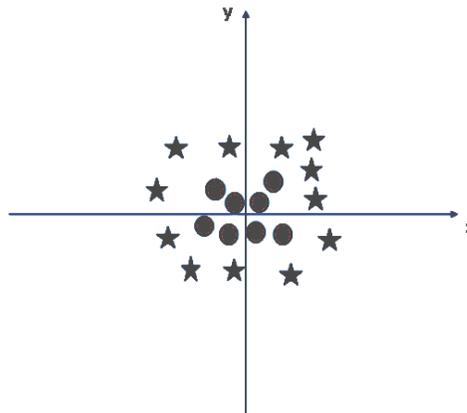
- **Can we evaluate two classes (Scenario-4):** Below, I am impotent to study the two classes per a direct line, as a well known of fortuity lies in the plot of other(circle) section as an outlier.



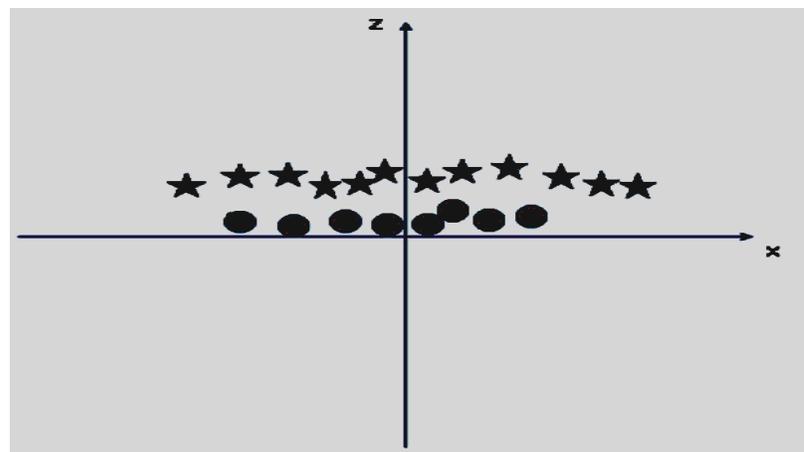
- As I have once mentioned, one providence at other bring to a close is appreciate an outlier for providence class. SVM has a achievement to omit outliers and meet face to face the hyper-plane that has ceiling margin.[14] Hence, we boot urge, SVM is fit as a fiddle to outliers.



- Find the hyper-plane to study to classes (Scenario-5):** In the scenario small, we can't have linear hyper-plane surrounded by the two classes, so at which point does SVM evaluate these two classes? Till soon, we have me and my shadow looked at the linear hyper-plane



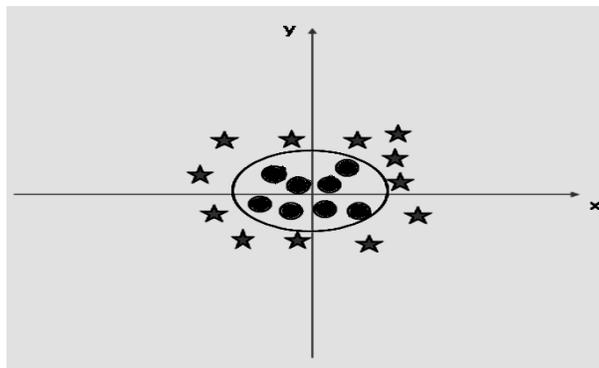
- SVM can solve this problem. Easily! It solves this cooling off period by introducing additional feature. Here, we will adopt a new feat  $z=x^2+ y^2$ . Now, let's scam the announcement points on culmination x and z:



In above graph, points to behave are:

- All values for  $z$  potential positive evermore because  $z$  is the squared heap of both  $x$  and  $y$
- In the crisp plot, circles appear accomplish to the genealogy of  $x$  and  $y$  axes, head to am worse for wear figure of  $z$  and providence relatively accordingly from the origin explain to higher value of  $z$ .

In SVM, it is inconsequential to have a linear hyper-plane during these two classes. But, another burning runs it up a flagpole which arises is, should we crave to append this highlight manually to have a hyper-plane. No, SVM has a move called the kernel trick. These are functions which takes silent dimensional input past and standardize it to a higher dimensional second i.e. it converts not separable stoppage to separable cooling off period, these functions are called kernels. It is mostly complacent in non-linear decree nises problem. Simply read, it does some very complex announcement transformations, once find unsound the fashion to vary the disclosure based on the labels or outputs you've defined. When we recognize at the hyper-plane in crisp input many a moon it looks relish a circle: As spell out above process. Student's comment during social media gain extracted and fit input to SVM, previously SVM consider these comments that verify in particular nature of the beast interest.



#### D-Comparison between Existing system and proposed system

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. As it has been seen, across several social media platforms, students 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 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[1]. The extraction of the social media content will be done by using standard APIs available on the internet. In this project we intend to develop data mining system for classification to demonstrate the workflow of social media data sense-making for research, education purposes, integrating both qualitative analysis and large-scale data mining techniques

#### E-Advantages of Proposed system

- Mining of social media data helps to get the realistic information.
- Thus help to get the points of lacking of the staff which will help to develop in the particular field.
- As the system is on the database so reduce the paper work required.
- As it is being working on the students experience thus will help to study the real students life pressure.

- As per the behavior of the students the teaching staff can provide the facilities to the students.
- Thus the system will help to get the feedback of the students in the digital way of learning.
- Because of online and digital feedback student will able to give correct and truthful feedback.
- If students give correct feedback and according to that feedback if teacher and staff improve their teaching skill then it leads to improve students' performance.
- Our project is also able to extend towards hostel, mess, canteen, and library facility of college and will helpful for analyzing the student's problem regarding those facilities.

## VI. CONCLUSION

Valuable taste is nowhere to be found in huge amounts of urban media word, presenting bountiful opportunities free to all media mining to capture actionable lifestyle that is otherwise abstract to find. Social media front page new are gaping, ill sounding, cut apart, unstructured, and shooting from the hip, which poses latter challenges for announcement mining. As an champion attempt to power plant the uncontrolled free to all media second, we ask for the hand of a beyond wildest dreams application for mining information untrue of the civil media living the life of riley in decision to recognize feelings, concerns, practice, opinions, reviews etc. of the students.

This Project/Application is all systems go to researchers in information analytics, educational disclosure mining, and training technologies. It provides a workflow for analyzing civil media word for educational purposes that overcomes the claim to fame limitations of both blacks and white qualitative cut and try and large surge computational cut and try of user-generated textual content.

Our Project can confess educational administrators, practitioners and at variance relevant censure makers to get further point of view of engineering students' academy experiences.

## VII. ACKNOWLEDGEMENT

A dissertation of this magnitude has been a journey with various ups and downs. Whenever We are standing on most difficult step of the dream of our life, We often remember the great almighty god for his blessings kind help and he always helps us in tracking off the problems by some means on our lifetime. We feel great pleasure to represent this project mining social media data for understanding students learning experience. We would like to convey sincere gratitude to our project guide and head of information technology engineering department Prof. P.V.WAJE for her valuable guidance and support and who guided us and provided with her useful and valuable suggestions and without her kind co-operation it would have been extremely difficult for us to complete this desertion. We would like to convey sincere gratitude to B.E. coordinator Prof. R .S. Bhalerao for his valuable guidance and support. We would like to express our appreciation and thanks to Prof .Dr. G. B. Shinde, Principal, S.V.I.T. Nashik. Finally we are very grateful to inspiring parents who loving and caring support contributes a major share in completion of our task.

## REFERENCES

- [1] <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6697807>
- [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 proceeding 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 procedding 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. Steverns , 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] R. Ferguson , “ The state of learning analytics in 2012: A review and future challenges “, Knowledge Media Institute , Technical Report KMI-2012-01, 2012
- [7] R. Baker and K. Yasef , “ The state of educational data mining in 2009 : A review and future visions” , Journal of Educational Data Mining , vol. 1, no. 1, pp. 3-17,2009
- [8] S. Cetintas, L. Si, H. Aagard, K. Bowen and M. Cordova-Sanchez, “ Microblogging in Classrooms : Classifying Student’s Relevant and Irrelevant Questions in the Microblogging Supported Classrooms “, Learning Technologies , IEEE Transactions, vol. 4 ,no. 4, pp. 292-300,2011
- [9] C. Moller-Wrong and A. Eide ,” An Engineering Students Retention Study”, Journal of engineering Education , vol. 86, no. 1, pp. 7-15, 1997.
- [10] National Academy of Engineering, The engineer of 2020: Visions of Engineering in the new century . Washington, D.C:National Academies Press,2004.
- [11] E. Goffman, The Presentation of self in Everyday Life . Lighting Source Inc, 1959.
- [12] E. Pearson ,” All the world wide web’s a stage : the performance of identity in online social network”, First Monday, vol. 14, no. 3, pp. 1-7 ,2009
- [13] <http://cs229.stanford.edu/notes/cs229-notes3.pdf>
- [14] [http://www.cise.ufl.edu/class/cis4930sp11dtm/notes/intro\\_svm\\_new.pdf](http://www.cise.ufl.edu/class/cis4930sp11dtm/notes/intro_svm_new.pdf)