Scientific Journal Impact Factor (SJIF) : 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745 ISSN (PRINT) : 2393 - 8161

Analysis Of Student OnlineActivities During Exams

A.Haritha[1], M.Jayanthi[2], K. Sri Vijaya[3],PVS Lakshmi[4]

[1]Asst Prof ,Dept of IT, PVP Siddhartha Institute of Technology &Research Scholar, Dept CSE, KLU,
[2]Student, Dept of IT, PVPSIT.[3]Asst Prof,, Dept of IT, PVPSIT.
[4]Professor Dept of IT, PVPSIT

Abstract— The indispensable expansion of social media and world wide web led to a huge growth in the number of active users and the volume of data it created on the online social networks is massive. We have huge volume of opinioned data available on the web we have to mine it so that we could get some interesting results out of it with could enhance the decision making process. Twitter is the most popular micro blogging service today. Many of the people are using twitter services like post short messages and sharing photos etc. So most of the researches are doing their analysis on twitter data. Here we are doing some different computational analysis which shows the student behavior at midterm exams, and also calculate the length of the tweets which are posted in exams time with some temporal granularity.

Keywords—extracting tweets, temporal granularity, length of tweets, classification algorithms, comparison, association rules.

I. INTRODUCTION

Twitter, founded in 2006, is the most popular micro blogging service today, with millions of its users posting short messages (tweets) every day [1]. This huge amount of user-generated content contains rich factual and subjective information ideal for computational analysis. In general, these massages could be classified into two groups: about Twitter users themselves and information sharing. Current research findings suggest that Twitter data could be utilized to gain accurate public sentiment on various topics and events. With help of some popular web scrappers, we collected tweets on the subject of midterm exams from students on twitter. Our aim was to investigate a real time Twitter sentiment on mid exams with some time limit. At different levels of temporal granularity, our analysis revealed the variation of different sentiment. Here what we want to observe is at which time the sentiment was too high and too low. We observed some consistent group behavior of Twitter users based on seemingly random behavior of each individual. And also we want to know at which time the tweets are increased than normal one, the tweets are decreased. Also we find out the length of the different tweets. Twitter users with positive sentiment appeared to have more friends and followers than those carrying negative sentiment. Also, users who shared the same sentiment inclined to have similar ratios of friends and followers, which is not true for general users.

II. METHODOLOGY

Our analysis is to know the behavior of the student towards exams. Here we are collected the tweets towards exams by using some popular web scrappers. After collecting the tweets we should clean the data then perform calculation to count the word frequency for that tweets which are gathered towards exams during on exams, before exams and after exams. We are creating some targeting key words which are related towards exams. Apply association rules to that key words then calculate the tweets scoring. Finally we get the graphical representation of the student behavior towards exams.

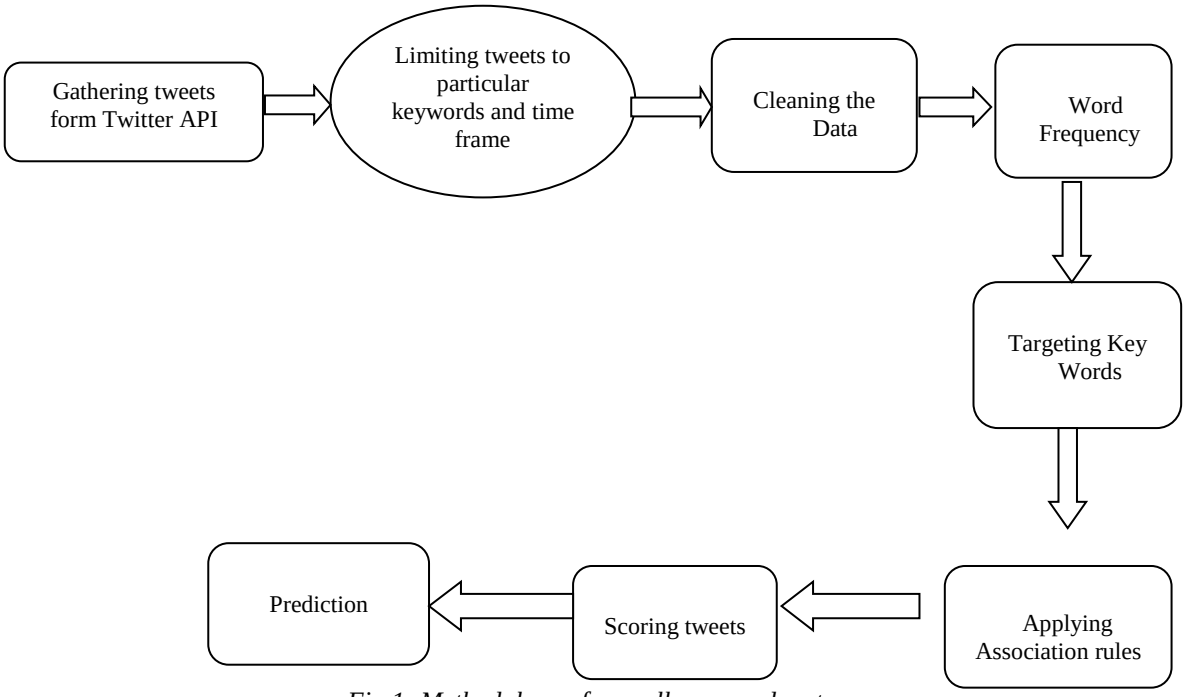


Fig 1: Methodology of overall proposed system.

2.1 Gathering tweets form Twitter API

Data extraction is can be done by using the web scrappers. Web Scrapping is a technique employed to extract large amounts of data from websites. Data from third party websites in the Internet can normally be viewed only using a web browser. Most websites do not offer the functionality to save a copy of the data which they display to your local storage[3]. The only option then is to manually copy and paste the data displayed by the website in your browser to a local file in your computer. With the help of the web scrapping we collect the tweets from twitter from the three different week's one on the exam, second before exams week, third after the exams week. At the various level of temporal granularity the tweets and time at which tweets are placed are collected. The extracted data is saved in different formats that can be excel, comma separated value.

2.2 Limiting the tweets

Due to the 140 character limitation on tweets, Twitters have adopted abbreviated and slang expressions to overcome this limit[4]. Tweets also contain misspellings, and are shorter and more ambiguous than other sentiment data such as reviews and blogs. Another feature of tweets is that they cover a variety of topics unlike other blogging sites that are more focused on one or a few topics[5]. As we mainly focus on tweet based on exam some of the words has been limited by specifying the size. There are several tools to limit the tweets based on the length or the text. Our main interest is to discover the fluctuation in the length about midterms from the particular group of twitter users by hour, day, and week. And also we want to find out the difference between the tweets which are gathered on exams, before exams, and after exams.

2.3 Clean the data

When we gather the tweets, the tweets may be in short cuts or it may contain some spelling mistakes and they have some unnecessary spaces. Before performing our analysis we must clean the data means we should remove all spaces and unnecessary data such as symbols, other information which is not related to perform our analysis. Today there are so many tools to clean the data, and also convert the given information as we need.

2.4 Word Frequency

Our main interest is to find out the length of the tweet, to know the length of the tweet first we should know the word count. And also we want word frequency to compare the tweets in different temporal granularity. Based on the word frequency we can easily differentiate the tweets which are posted in exams time and the normal time. Based on this analysis we can easily analyze, how the student activities are in exams and in remaining time. There are so many popular tools are available to calculate the word count.

2.5 Targeting Key Words

After calculating the word count we are specifying some key words which are most related to our analysis. These key words are called as targeting key words. Here we are not considering all the words which specified in the word count, as given above only some words which are used to create some association rules which are useful to our analysis. Means here we are taking the targeting key words are related to exams like time, exam, stress, etc. to perform the analysis[6].

2.6 Applying Association Rules

To create association rules we need some frequent item sets. Here we can generate frequent item sets by using FP-Growth algorithm. In FP-Growth algorithm to generate frequent item sets we can taking targeting key words. In FP-Growth algorithm to generate frequent item sets we are calculating support and confidence to the given target words.

Support(s) of an association rule is defined as the percentage/fraction of records that contain X ∪Y to the total number of records in the database. [2]

Support (XY) = Support count of XY / Total number of transaction in D

Confidence of an association rule is defined as the percentage/fraction of the number of transactions that contain X ∪Y to the total number of records that contain X, where if the percentage exceeds the threshold of confidence an interesting association rule X ⇒Y can be generated. [2]

Confidence (X/Y) = Support (XY) / Support(X)

Based on FP-Growth algorithm here we are generating some association rules which are related towards exams. By selecting one of the rule in given several rules we can get the output that is in exams time and rest of the student online activities.

2.7 Scoring Tweets

In this we are calculating the length of the tweets by using some classification algorithms. First we can save the tweets as .CSV (Comma Separated Value) file. Then we can import the file then apply the classification algorithm to that file. Based on our interest we can apply any classification algorithm, here we are using naïve Bayes algorithm. We can any parameter in our analysis only two parameters are there such as time and tweets. Based on these we can get the length of the tweet[7].

2.8 Analysis

Data extraction is can be done by using the web scrappers. Simplehtmldom is a PHP library that is used for creating web scrappers[8]. The data extracted are placed in .txt files. By using naïve Bayes algorithm as we specified before at what time the tweets have been occurred in maximum length or minimum length that is after exam, before exam, and on exams. Based on calculating the support and confidence we are analyzing the different frequent item sets that are related towards exams during on exams, before exams and after exams.

III. RESULTS AND DISCUSSIONS

Fig 2: Extracting tweets during midterm exams.

Here we are extracting tweets at exams time by using some web scrappers. We are extracting any type of data form web scrappers by specifying the website name. After giving the web site name the web page will be open then we should give our details for the authentication purpose. Then we enter into our page and extract the data.

Here we can store the tweets which can extracted like on exams,before exams and after exams. Means this is a one type of database. The tweets we are extracted in different temporal granularity which are saved in comma seperated value(CSV) files for calculating length of the tweets in different timings.We can retrieve the CSV files which are stored in the repository. First we can drag the on exams tweets file into main screen. Then apply the naïve bayes aalgorithm to that file by using some label like we can take label as a tweet or else we can take time. Based on the label the length will calculated. We are applying the same process on three different types of tweets such as on exams, before exams and after exams.

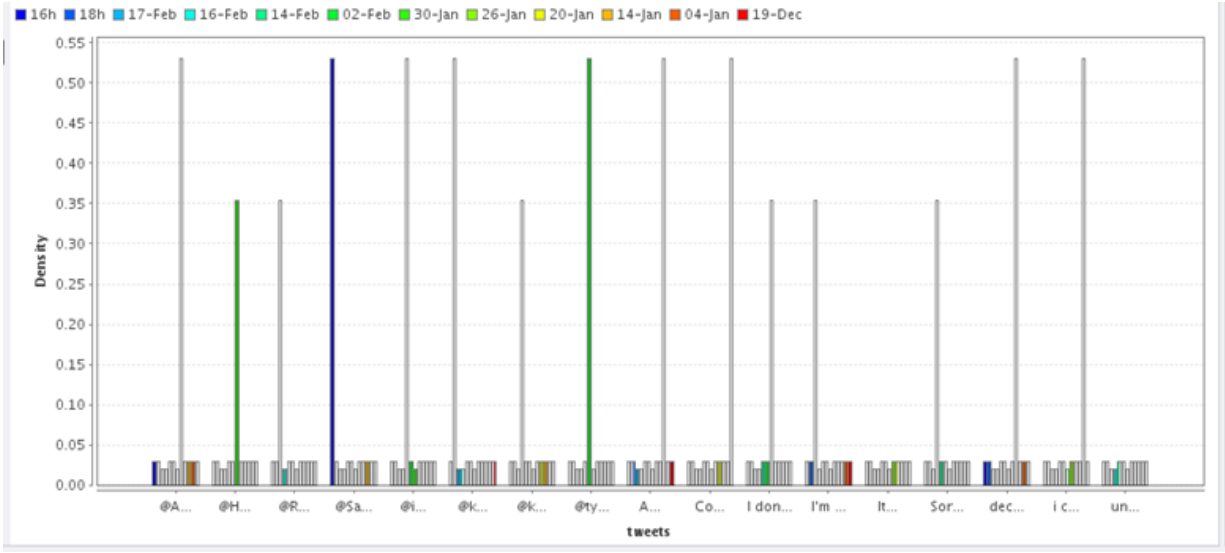


Fig 3: Length of the tweets which are treated as on exams tweets.

Here we are considering the density as length of the tweets which are taking in different temporal granularity in exams time. And the colours here specified are showing the different dates of tweets. Same procedure as before says in exams time tweets here also we can do same thing for calculating the length of before exams time tweets. But only the difference is here we are retrieve the tweets which are gathering in before exams time form the repository.

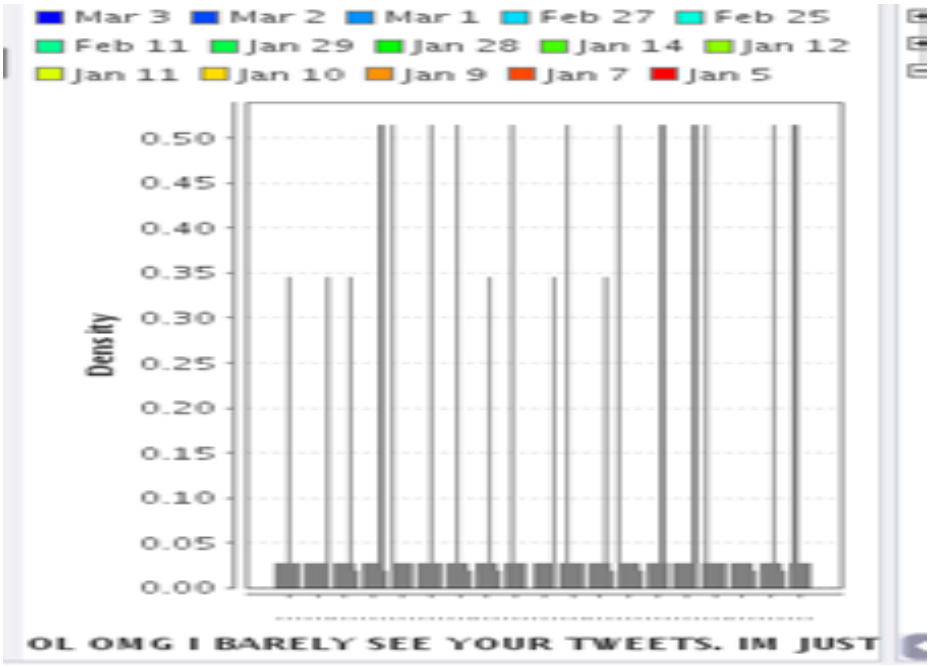


Fig 4: Length of the tweets which are in before exams time.

Same procedure as earlier explained in exams time tweets length calculation but only the difference is here we can retrieve the tweets which we are extracting after exams time from the repository.

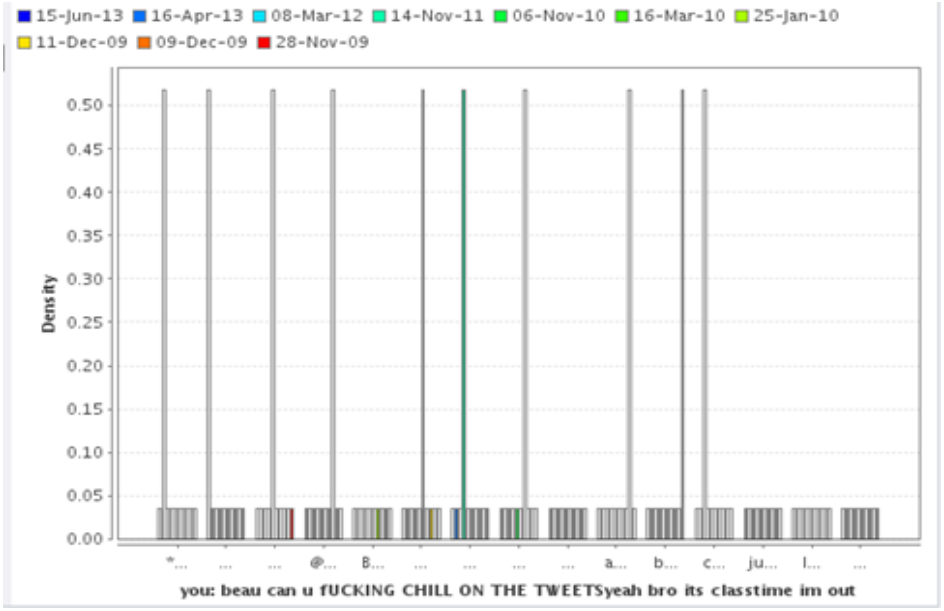


Fig 5: Length of the tweets which are in after exams time.

By calculating the length of the tweets which are gathered in different exam timings such as on exams, before exams and after exams in different temporal granularity our analysis is completed half of the part.

The next part is comparing the tweets which are taking in three different exam timings such as on exams, before exams and after exams. By comparing the tweets we can analyze the student behavior in exams time and the rest of the time. Means the feelings of the students in exams time like he or she may feel more stress than normal time or not in this way we know feelings of the student.

Here we should import the three types of files such as on exams, before exams and after exams. For calculating the length of the tweets we can save the tweets as comma separated value (CSV) files but here we are comparing the tweets so we can save these three files in note pad. And these three are imported in a single document.

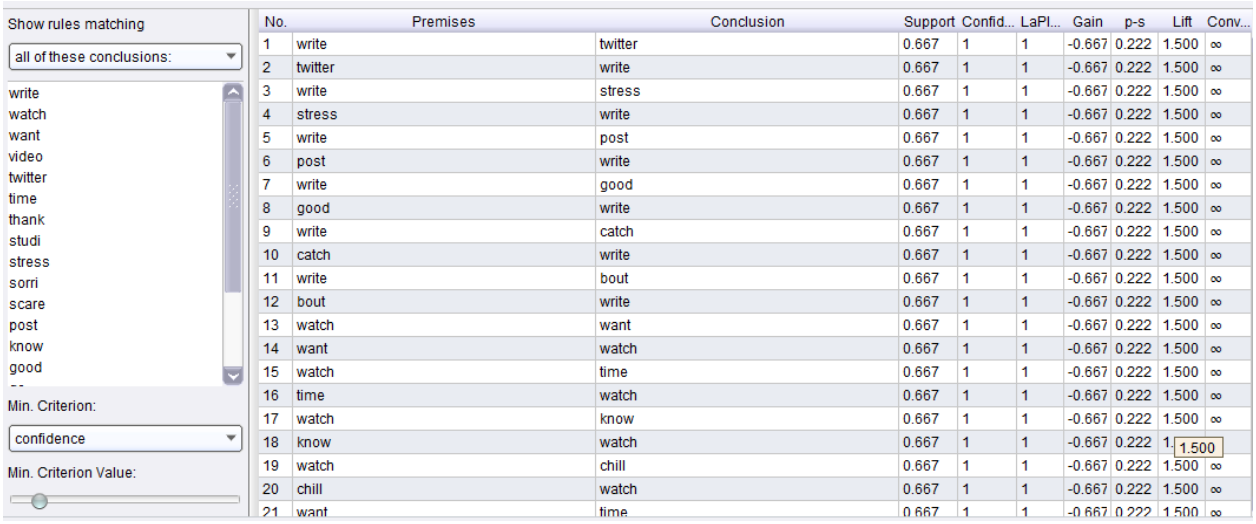
Here we are apply tokenize for dividing the tweets into different tokens. And then apply filter stop words, filter tokens by length, stem porter, transform cases algorithms in the document in which the three files are placed. Why we are applying all these algorithms, to compare the tweets we need some keywords which related to the exam. So here the spaces, stop words, divide the tweets into different stems, we can specify the minimum and maximum length of the by using filter token by length algorithm and finally the transform cases are used to convert all the text into single case either in upper case or lower case.

Word	Attribute Name	Total Occurences	Document Occurences	tweet	tweetafterexam	tweetbeforeexam
aanandapadataniki	aanandapadataniki	1	1	0	1	0
abl	abl	1	1	1	0	0
activ	activ	1	1	1	0	0
actual	actual	1	1	0	1	0
advis	advis	1	1	0	1	0
alexpistoriu	alexpistoriu	2	1	2	0	0
angel	angel	1	1	0	0	1
anger	anger	1	1	0	1	0
annoi	annoi	1	1	1	0	0
antha	antha	1	1	0	1	0
arr	arr	1	1	1	0	0
arvindkejriw	arvindkejriw	1	1	0	0	1
asleep	asleep	1	1	0	1	0
assign	assign	1	1	0	0	1
assum	assum	1	1	0	0	1
avail	avail	1	1	1	0	0
aypoyai	aypoyai	1	1	0	1	0
babi	babi	1	1	0	1	0
baga	baga	1	1	0	1	0
bare	bare	1	1	0	0	1
bdai	bdai	1	1	0	0	1

Fig 6: Word lists of three files.

After applying all the stemming and filtering algorithms finally we get this word list[9]. Form those three files which contains the tweets on exams, before exams and after exams. In those three types of files the words are taken which are related to the exam and the behavior of student. The word list contains the words in three files how many types they are repeated in each file and all the three files.

To know the behavior of the student we are applying some association rules or we can create association rules. To create association rules we are using FP-Growth algorithm because the association rules are created by using frequent item sets. The FP-Growth algorithm is not taking text document directly for generating frequent item sets. So first we can convert the text into numerical and then again convert the numerical to binomial because the FP-Growth algorithm takes only binomial data for generating frequent item sets.



No.	Premises	Conclusion	Support	Confid...	LaPl...	Gain	p-s	Lift	Conv...
1	write	twitter	0.667	1	1	-0.667	0.222	1.500	∞
2	twitter	write	0.667	1	1	-0.667	0.222	1.500	∞
3	write	stress	0.667	1	1	-0.667	0.222	1.500	∞
4	stress	write	0.667	1	1	-0.667	0.222	1.500	∞
5	write	post	0.667	1	1	-0.667	0.222	1.500	∞
6	post	write	0.667	1	1	-0.667	0.222	1.500	∞
7	write	good	0.667	1	1	-0.667	0.222	1.500	∞
8	good	write	0.667	1	1	-0.667	0.222	1.500	∞
9	write	catch	0.667	1	1	-0.667	0.222	1.500	∞
10	catch	write	0.667	1	1	-0.667	0.222	1.500	∞
11	write	bout	0.667	1	1	-0.667	0.222	1.500	∞
12	bout	write	0.667	1	1	-0.667	0.222	1.500	∞
13	watch	want	0.667	1	1	-0.667	0.222	1.500	∞
14	want	watch	0.667	1	1	-0.667	0.222	1.500	∞
15	watch	time	0.667	1	1	-0.667	0.222	1.500	∞
16	time	watch	0.667	1	1	-0.667	0.222	1.500	∞
17	watch	know	0.667	1	1	-0.667	0.222	1.500	∞
18	know	watch	0.667	1	1	-0.667	0.222	1.	1.500
19	watch	chill	0.667	1	1	-0.667	0.222	1.500	∞
20	chill	watch	0.667	1	1	-0.667	0.222	1.500	∞
21	want	time	0.667	1	1	-0.667	0.222	1.500	∞

Fig 7: Association rules are generated to know the student behavior.

The association rules are generated by using frequent item sets. Based on the rules we know the student behavior in exams time[10,11]. From all these rules each and every rule give us one type of student behavior. For example we can take time as a rule we know the student in the particular time he or she may enjoy the exams or boring or feeling more stress etc. based on the rule the behavior of the student may change.

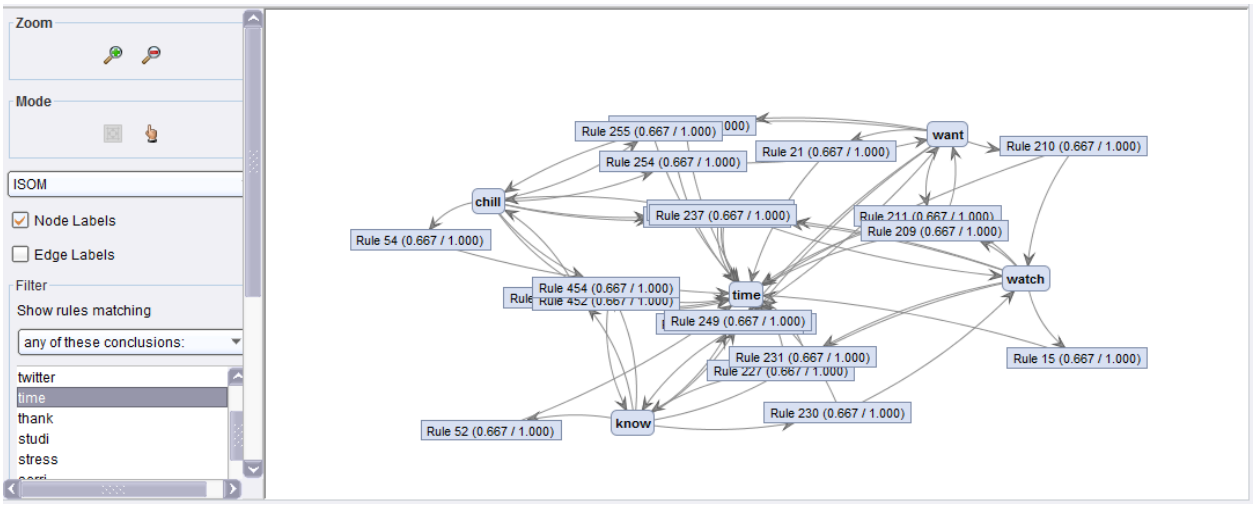


Fig 8: Association of word time with respect to other words.

IV. CONCLUSION

As the world is getting updated the derogatory of the student has been changing in every aspect in our daily routines. Irrespective of the importance of the issue, students make their own note in the social blogs. The proposed system retrieves the tweets posted in social blog mainly on exams time and specify the length of the tweet at particular time granularity[12] and also know the behavior of the student in exams time. In opinion mining many of people are think about the sentiment analysis but we can do different types of analysis. In this we are analyzing student behavior in mid exams and also calculating the length of the tweets which are posted in different temporal granularity. We performing this analysis by using some classification and association rules[13]. This system is driven with a motivation that, it has to be easily understood and can be analyzed by any novice user to an analytical expert. The graphical representation gives crystal clear prediction.

REFERENCES

[1] Wei Hu, Department of Computer Science, Houghton College, New York, USA, Real-Time Twitter Sentiment toward Midterm Exams, Received December 3rd, 2011; revised January 7th, 2012; accepted February 6th, 2012.
[2] QiankunZhao,Nanyang Technological University, Singapore and Sourav S. Bhowmick, Nanyang Technological University, Singapore.Association Rule Mining: A Survey.
[3] Harb, A. , Plantié, M. , Dray, G. , Roche, M. , Trousset, F. and Poncelet, P. , "Web Opinion Mining: How to extract opinions from blogs?", International Conference on Soft Computing as Tran disciplinary Science in 2008.
[4] Krishnamurthy, B., Gill, P., & Arlitt, M. (2008). A few chirps about twitter. Proceedings of the First Workshop on Online Social Net- works, Seattle, 17-22 August 2008, 19-24.
[5] A. BALAHUR and A. MONTOYO, "A Feature Dependent Method for Opinion Mining and Classification" Natural Language Processing and Knowledge Engineering, 2008. NLP-KE '08. International Conference on Date of Conference: 19-22 Oct. 2008.
[6] X. WANG, G. HONG FU "Chinese subjectivity detection using a sentiment density-based naive Bayesian classifier" conference on machine learning and cybernetics (icmlc) international in 2010.
[7] Rillof, E. , Wiebe, J. , and Philips, W. , "Exploiting Subjectivity Classification to Improve Information Extraction", Proceedings of the 20th National Conference on Artificial Intelligence in 2005.
[8] N. M. Shelke, S. Deshpande and V. Thakre "Survey of Techniques for Opinion Mining" International Journal of Computer Applications (0975 – 8887) Volume 57– No. 13, November 2012.
[9] G.Angulakshmi1, Dr.R.ManickaChezian, " An Analysis on Opinion Mining : Techniques and Tools", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 7, July 2014, Pages 7483-87, 2014.
[10] Nidhi R. Sharma , Prof. Vidya D. Chitre, "Opinion Mining, Analysis and its Challenges" , International Journal of Innovations & Advancement in Computer Science, IJIACS, ISSN 2347 – 8616, Volume 3, Issue 1, April 2014.
[11] B. B. Greene and G. M. Rubin, "Automatic grammatical tagging of English," Technical Report, Department of Linguistics, Brown University, 1971.
[12] Sentiment Analysis and Opinion Mining Bing Liu. Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.
[13] Muldner, K., Burleson, W., Van de Sande, B., & Vanlehn, K. (2011). An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. User Modeling and User - Adapted Interaction, 21(1-2), 99-135. doi: 10.1007/s11257-010-9086-0.