



# **A Machine Learning Approach for Understanding Student Learning Experience Using Naive Bayes Classification**

V. Abinaya<sup>1</sup>, Dr. B. Srinivasan<sup>2</sup>

M.Phil Scholar, Dept of Computer Science, Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India<sup>1</sup>

Associate Professor, Dept of Computer Science, Gobi Arts & Science College, Gobichettipalayam, Tamilnadu, India<sup>2</sup>

**ABSTRACT:** Data mining is the research challenge of the "Studying Discovery in database" and retrieving information from the warehouse is not only tedious but also difficult in some cases. The existing algorithm does not provide fast computation and better result. The classification assignment is classified with the aid of the well-described lessons, and a training set along with reclassified examples and Naive Bayes algorithm has been used to predict students learning experience from social network. Such as Twitter, Face Book and Tumblr. Naive Bayes is a classification algorithm for binary (elegance) and multi-magnificence type issues. These machine has been used for the supervised gaining knowledge of method for the corporation of categorized the feedback statistics on social media and Naive Bayes algorithm is classified into the scholars records set.

**KEYWORDS:** Educational data mining, Naive Bayes Classifier, weka tool, social media.

## **I. INTRODUCTION**

Data mining is broadly used to classification technique and it includes examining the capabilities of newly provided item and assigning to it a predefined class. The classification assignment is classified with the aid of the well-described lessons, and a training set along with reclassified examples and Naive Bayes algorithm has been used to predict students learning experience from social network. Such as Twitter, Face Book and Tumblr. Therefore, in the proposed system using Naive Bayes is a classification algorithm for binary (elegance) and multi-magnificence type issues. These machine has been used for the supervised gaining knowledge of method for the corporation of categorized the feedback statistics on social media and Naive Bayes algorithm is classified into the scholars records set. Proposed methodology builds this using mat lab framework. The working flow of Naive Bayes classification algorithm on matlab using weka tool and semantic student's datasets are used for the experimental analysis.

## **II. LITERATURE REVIEW**

AJINKYA.A.LONDHE, JAMGEKAR R.S, SOLUNKE B.R. has proposed to college students getting to know revel in through social media. Studies from the fields of information mining have excellently produced diverse technique, tools, and algorithms for management massive quantities of data to reply actual-global issue. Social media is normally used for an expansion of purposes; sizable quantities of user created information can be made existing for statistics mining. Social network is for all and sundry and it is now the type of large and huge a part of our complete lives. In step with the survey contribution of the information is high in the social websites like twitter and face e book.

G.SIEMENS AND P.LONG has proposed to analytics in studying and gaining knowledge of and it affords an incredible platform for college students to bring their views, pressure, feelings, critiques, troubles, pleasure, war, opinion and searching for social support. Scholar's speak and share their ordinary encounters in formal and informal way on exceptional social media sites. This pupil's tweets and submit can present valuable and implicit records that is probably very beneficial for an institution to understand the problems of the student he/she facing in the mastering gadget. Hence, enhancing education exceptional, and thus decorate scholar employment, retention, and fulfillment.



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 2, February 2018

M. ROST, L. BARKHUUS, H. CRAMER, AND B. BROWN, has describes about college students getting to know experience in social media interpretation. Those social network records mining offer a scope to regulate training machine as college students play an important function sooner or later to make an impact on country's financial boom.

MS. PRANJALI S. JADHAV, DR. SHIRISH S. SANE have needed-for a new solution for student getting to know revel in. The computer and mastering schooling targets to increase expertise and knowledge of manner wherein virtual technology can enhance schooling. Considering that to recognize, The Social Media all through 21th century may be very properly-liked the Twitter and Face book are generally used. They shared their academic revel in which includes their thoughts, critiques, their emotions and new matters approximately getting to know procedure. Such instrumented environments will provide is powerful for the data of student learning. The pupil's excellent expertise is reflected in social media sites which desire human interpretation.

MS. PRANJALI S. JADHAV, DR. SHIRISH S. SANE has proposed a new answer for scholar gaining knowledge of revel in. The up-and-coming field of schooling analytics and studying data mining has decided on reading based facts received from course management systems (CMS), school room era utilization, or managed on line getting to know environments to inform educational selection making. however, to the pleasant of our information, there is no studies located to immediately mine and examine scholar published content material from out of control areas on the social internet with the clear goal of expertise pupil's learning stories.

PAYAL S.JAIN, PALLAVI S.PANHALE, PRANESHWARI A.DEOKAR has described with reference to treasured mining in social media records for college students mastering experience. Several times, pupil's gets shy or trepidations of clearing their catch twenty two situation the classroom and this convivial media avail them to just post something they sense at that point about their emotions. The faculties and departments have been struggling with pupil recruitment and withholding troubles.

### III. PROPOSED ALGORITHM

Mining the social media information like engineering students observe problem will end result to categorize the collection of Engineering scholar's consistent with their reports and become aware of their issues to be solved to enhance the education great. The facts collection is made directly mine and analyses scholar-posted content material with thinking about the Engineering students problem from uninhibited spaces on the social web with the clear aim of expertise engineering students gaining knowledge of reviews. The present paintings has now not measured pupil academic usual overall performance to turn out to be privy to the Engineering university students problem and classify them because it ought to be for enhancing E-studying reports.

Doing this, the proposed gadget is identifying and classifying the strain in completes their troubles confronted by engineering college students to improve their schooling best with admire to their properly and advantageous comments.

### NAIVE BAYES CLASSIFICATION

Naive Bayes version is straightforward to construct and especially helpful for extremely massive facts units. Together with simplicity, Naive Bayes is thought to higher even rather complicated class strategies.

Bayes theorem provides a way of calculate posterior probability  $P(c|x)$  from  $P(c)$ ,  $P(x)$  and  $P(x|c)$ .

Look at the equation below:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Above,

- $P(c|x)$  is the posterior probability of class (c, target) given predictor (x, attributes).
- $P(c)$  is the prior probability of class.
- $P(x|c)$  is the possibility which is the probability of predictor given class.
- $P(x)$  is the prior probability of predictor.

### NAIVE BAYES CLASSIFICATION ON SENTIMENTAL AND LEARNING ANALYSIS

Naive Bayes multi-level classifier is used for classification of reports by locating the opportunity of phrases in tweet and post for every class possibility of every label includes how many users. Ultimately the tweets with new Label could be compared with the relaxation of the tweets with present Labels.



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 2, February 2018

The Naive Bayes Classifier approach has been based totally the use of Bayesian theorem and the high-quality nicely-matched to the dimensionality of the inputs has higher result. Again and again once more outperforms extra complicated class methods, but it is straightforward to activate. For the models, maximum probability estimates all the parameters. To estimate the parameters, it has been contain of small range of education. It operates glowing and efficaciously in supervised studying.

## In Bayesian Evaluation,

- **Prior chance:** its miles a precept and based totally on previous enjoy. it is a ratio of wide variety of on its personal items and number of entire objects.
- **Likelihood:** To categorize an authentic item that this item belongs to which case.
- **Posterior probability:** The complete elegance is absolute thru combining both bases of data i.e. previous and chance to form a Posterior opportunity the usage of Bayes rule.

The Posterior probability of X man or woman an object  $\alpha$  earlier opportunity of total items likelihood of X given gadgets.

Above theorem guide us to resolve the conditional possibility of contrary and unbiased occasions. The estimation of the possibility for an opinion might be containing high quality, bad or impartial. It might produce appropriate consequences.

The Naive Bayes entire glowing in higher dependent features and outperforms often compared to different text association algorithms. Despite the fact that criterion particularly risk parameter getting to know for Naive Bayes classifier has a tendency to be suboptimal.

## IV. EXPERIMENTAL SETUP

### MODULE DESCRIPTION

#### 1. TEXT PRE-PROCESSING

Twitter users use some unique symbols to deliver confident which means for instance, # is used to suggest a hash tag, @ is used to indicate a user account, and RT is used to signify a re-tweets. Twitter customers once in a while reiterate letters inwards in order that to highlight the phrases, as an example, huuungryyy?,soomuuchh?, and Monnndayyy?. Besides, ordinary forestall phrases along with a, an, and, of, he, she, itl, non-letter symbols, and punctuation also deliver noise to the textual content. So we pre-processed the texts in advance than education the classifier.

- To eliminated all the #research scholar troubles hash tags. For other co-taking place hash tags, and best eliminated the #sign, and saved the hash tag texts.
- Bad words are beneficial for detecting bad study trouble and troubles. So that substituted phrases ending with NT (poor token) and different common negative phrases (e.g., not anything, never, none, cannot) with poor token.
- To remove all words that incorporate non-letter code and punctuation. This integrated the elimination of @ and http hyperlinks, and also disconnected all of the RTs. For repeating letters in phrases, our method turned into that when to detected identical letters repeating, stored both of them. If were detected more than two identical letters repeating, and changed them with one letter.
- Consequently, hungry and so were corrected to hungryl and sol. Lots were stored as a good deal. At first correct phrases such as tool and sleep| had been kept as they were. We used the Lemur statistics retrieval toolkit to dispose of the common prevent phrases. To save words like lots, extra, all, always, still, simplestl, because the tweets often use those words to express extent. The caretaker stemmer inside the Lemur toolkit changed into used to execute stemming as a way to unify changed sorts of a word, including plurals and modified styles of a verb.

#### 2. STEMMING

Stemming is the expressions used in linguistic morphology and records revival to offer an explanation for the method for plummeting inflected (or on occasion derived) phrases to their phrase stem, base or root shape generally a written word form. The stem desires not to be matching to the morphological root of the word; it is also quality that



# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 2, February 2018

associated words map to the identical stem, even if this stem is not in itself a legitimate root. This module is used to suggestion of five labels and “Others” label as more case to fetch pointless tweets.

## Key point Localization

A key point has been discovered with the aid of examine Labels to its neighbors and is to execute a detailed suit to the close by facts for vicinity, scale, and ratio of key phrases. The low evaluation points or poorly localized along a labellers are eliminated with the aid of key factor localization. Records similarity evaluation is essentially essential to numerous multimedia information processing systems and programs, inclusive of pre-processing, restoration, label enhancement, replica detection, retrieval, and naive bayes rule popularity/classification. The principle aim of facts similarity attention is to layout algorithms for mechanical and goal assessment of similarity in a way that is regular with subjective human question reviews.

## Descriptor Generation Using Vocabulary Tree

The local vicinity descriptors are hierarchically quantized right into a vocabulary tree. The vocabulary tree permits a bigger and greater discriminatory vocabulary for use efficiently, which leads to a dramatic improvement in retrieval quality. The maximum extensive belongings of the scheme are that the tree immediately defines the quantization. The quantization and the indexing are fully included.

## Bow (BAG of Words) Representation

Bow illustration approach quantizes dataset labels feature descriptor by means of naive classification into a collection of visual phrases pedestal on the visual language tree. The Bow representation approach assessing the similarity between QAW feature descriptors may be measured by matching their corresponding visual words via label matching and it categorized on associated way.

## DATASET GENERATION

The data set used for this paper is inside the form of text files and to used textual content dataset that's collected by using personal. There are fifty discussions are completed. They were labeled manually by self. Labels belong to five distinctive category training, together with 'finance', 'work', 'know-how', 'state of affairs' and 'getting to know experience'.

The full range of categories is ten now not precisely recognized, but lots of them arise best very rarely. The dataset is split in twenty two posts of unmarried files delimited by means of text tags. The labels are categorized by using following bag of phrases.

- Lack of information (set of rules, problem, data literacy, records behavior, PhD students, humanities, general abilities)
- Lack of motivation (thoughts, conceive, emotions, temptation, pretty tough)
- Lack of expertise in equipment and software program (proposed software, ns2, weka, rapid miner, logic paintings, tough algorithm, factors, good judgment, software program problems)
- Paintings pressure (documentation, residence work, legacy, rearranging, file control,)
- Financial problems (cash, timings, feepostpone, hours, wealth, price, implementation and value)

Twenty five new PhD college students responded to the preliminary questionnaire. One of the respondents indicated that they felt 'very confident' approximately the literature search but nine (36%) indicated that they felt 'confident'. another nine (36%) had been 'somewhat confident', even as six (24%) have been 'not that assured' and one became 'never assured'. Fifteen (60%) claimed to be both 'absolutely decided' or 'close to completely decided' on their research subject matter.

Seven (28%) have been 'incredibly Ronan Madden 282 determined', and three (12%) were 'now not truly decided'. no one indicated that they were 'very aware' of the facts assets they were probable to use, however fifteen (60%) claimed they had 'an awesome concept' of the assets (20%) indicated they'd 'a few idea' of the resources, while another five have been either 'no longer too sure', or 'never certain'. whilst one student claimed to be 'very assured' in using the assets, eleven (forty four%) indicated they were 'assured'. Six (24%) were 'somewhat confident', and seven (28%) had been either 'a piece uncertain', or 'not at all sure' of using the assets.

# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 6, Issue 2, February 2018

## EXPERIMENTAL SETUP

Usually used measures to evaluate the overall performance of type models encompass accuracy, precision, and the harmonic common among precision and keep in mind and the F1 rating. For multi-label type, the occasions are to a point extra complicated, due to the fact each record gets assigned a couple of labels. Among these labels, a few may be accurate, and others may be wrong. Consequently, there are typically two forms of evaluation measures example based measures and label-primarily based measures. example based totally measures are calculated on every report (e.g. every tweet is an article, and additionally called an illustration right here) after which averaged greater than all documents inside the dataset, while label-primarily based measures are designed based totally on each label (category) after which averaged over all labels (classes).

It considers similarly the precision p and the remember r of the take a look at to work out the rating: p is the wide variety of correct superb effects divided by using the wide variety of all wonderful results, and r is the wide variety of accurate tremendous effects divided by way of the quantity of fantastic results that should have been again. The F1 rating can be interpreted as a weighted average of the precision and remember, in which an F1 score reaches its most wonderful price at one and worst at zero.

The standard F-measure or balanced F-score (F1 rating) is the harmonic mean of precision

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

The general formula for positive real  $\beta$  is:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

That is the probability of word in specific category is equal to words appears into that category divided by total number of word. For Ex: Here our comment is "hard much info class Wait No both hard extensive". First we calculate probability of word belongs to category such as "Heavy Study Load" the result of this is 1.0 And probability of word belongs to category other than "Heavy Study Load" the result of this is 1.0. The whole table for Heavy Study load is shown in Table 1.

C	C	C'	C'
	Probability		Probability
0.0	0.0	1	1.0
1.1	1.1	8	1.0
0.0	0.0	8	1.0
0.0	0.0	10	1.0
0.0	0.0	12	1.0
0.0	0.0	15	1.0
1.0	1.0	8	1.0
1.0	1.0	5	1.0
0.0	0.0	6	1.0

Table 1. Probability Calculation

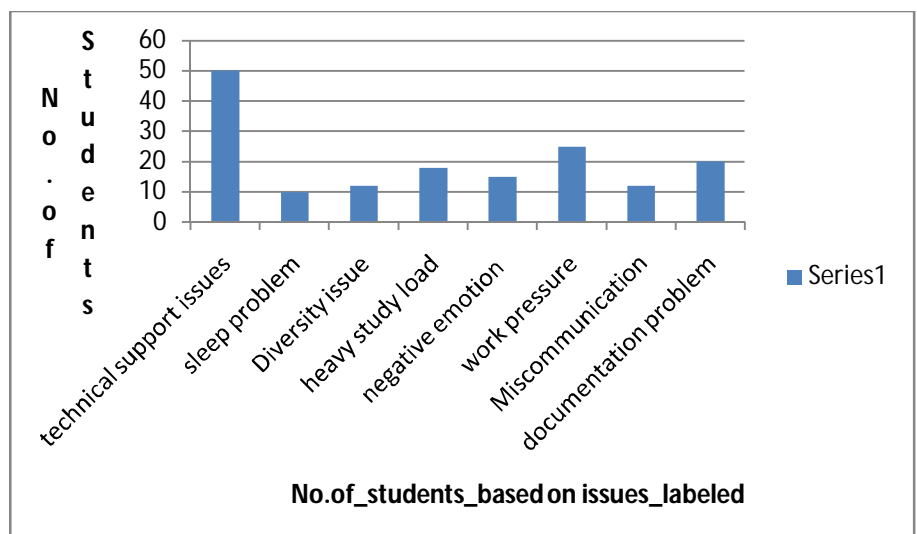


Fig. 2 Graphical Representation of Tweets Classification





# International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 2, February 2018

## V. CONCLUSION AND FUTURE WORK

A selection of social media websites of engineering scholars are discussed and contribute to their each day encounters in a casual behavior, reading such statistics but can be hard and the complexity of studies stage, human beings experiences reflect from social media content material requires human interpretation. However the increasing scale of data loads automatic statistics evaluation techniques. This work recognition on to reveal a workflow of social media data experience advent for instructional functions, integrating each qualitative evaluation and massive scale facts mining techniques and to find out research point human beings troubles come upon in studying process and this also gives the protection of Engineering students. The Naive Bayes type technique proved the proposed works efficiency and the results are showed as more accuracy. Future paintings may want to analyze education surroundings generated content material others than texts (e.g. data and videos), on social media sites aside from Twitter (e.g. Face e-book, Tumblr and YouTube). Future work can be executed to design greater state-of-the-art algorithms in order to expose the hidden records in the long tail. The manipulation of private records on line can also want to be taken into issues in destiny.

## REFERENCES

- [1] Arun K Pujari, "Data Mining Techniques", University Press, Education 2001.
- [2] Ajinkya A. Londhe, Jamekar R.S and Solunke B.R, "Analyzing Student's Learning Experiences Through Social Media Data Using Machine Learning Tool", N.B Navale Sinhgad College of Engineering, Solapur, International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835, Volume-3, Issue-12, Dec-2016.
- [3] Akanksha A. Pande, Kinariwala S.A, "Analysis of Student Learning Experience by Mining Social Media Data", International Journal of Advances in Electronics and Computer Science, Volume 7 Issue No.5, MIT College, Aurangabad, India.
- [4] Archana.T and Usha Devi Gandhi, "Prediction of Student Performance in Educational Data Mining", VIT University, Vellore, 2016.
- [5] Beck, J., & Woolf, B. (2000), "High-level student modeling with machine learning". Proceedings of the 5th International Conference on Intelligent Tutoring Systems, 584–593
- [6] Cristobal Romero (2010), "Educational Data Mining: A Review of the State-of-the-Art", IEEE Transactions on systems, man and cybernetics-Part C: Applications and Reviews vol. 40 issue 6, pp 601 – 618.
- [7] Dorina Kabakchieva (2012), "Student performance by using data mining classification Algorithms", International Journal of Computer Science and Management Research Issues Vol. 1, Issue 4, November 2012.
- [8] Dringus, L.P., Ellis, T. (2005), "Using data mining as a strategy for assessing asynchronous discussion forums", Computer and Education Journal, 45, 141–160.
- [9] Gauri Joshi and Mr. Samadhansonawane, "Filtering and Classification Of User Based On Social Media Data Using Memetic and Naive Bayes Methods", Savitri Bai Phule University of Pune SITRC, Nasik, India, International Journal of Computer Science and Information Technologies, Volume 6(5), 2015, 4795-4798.
- [10] Jaiwei Han, Micheline Kamber, "Data Mining: Concepts and Techniques".
- [11] Jiangtao Ren, Sau Dan Lee, Xianlu Chen, Ben Kao, Reynold Cheng and David Cheung, "Naive Bayes Classification of Uncertain Data", Sun Yat-sen University, Guangzhou, 510275, China.
- [12] Lalit Dole, Jayant Rajurkar, "A Decision Support System for Predicting Student Performance", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 12, December 2014.
- [13] Maneesh Singhal and Ramashankar Sharma, "Optimization of Naive Bayes Data Mining Classification Algorithm", International Journal For Research in Applied Science and Engineering Technology, Vol 2, Issue 8, August 2014, Rajasthan Technical University, Kota, Rajasthan, India.
- [15] Mukesh Kumar, Shankar Shambhu & Punam Aggarwal, "Recognition of Slow Learners Using Classification Data Mining Techniques", Imperial Journal of Interdisciplinary Research (IJIR) ISSN: 2454-1362 Vol-2, Issue-12, 2016.
- [16] Mrs. Mythili M.S, Dr. Mohamed Shanavas.A.R, "An Analysis of student's performance Using Classification algorithms", IOSR Journal of Computer Engineering, Volume 16, Issue 1, January 2014.
- [17] Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt-Thieme (2009), "Improving Academic Performance Prediction by Dealing with Class Imbalance", Ninth International Conference on Intelligent Systems Design and Applications
- [18] Dr. Nithya. P, Umamaheswari. B, Umadevi. A (2016), "A Survey on Educational Data Mining In Field of Education", Department of CS, PSG CAS, Coimbatore. Tamil Nadu, India. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 1, January 2016.
- [19] Pallavik. Pagare, department of computer engineering, "Analyzing Social Media Data for Understanding Student's Problem", International Journal of Computer Applications (0975-8887), MET's Institute of Engineering, Nashik, Savitri Bai Phule Pune University, Maharashtra India.
- [20] Parneet Kaur, Manpreet Singh, Gurpreet Singh Josan, "Classification and prediction based Data Mining algorithms to predict slow learners in education sector", Scholar, Department Of CSE, Punjab Technical University, Jalandhar 144603, India, Assistant Professor, Department CSE&IT, GNDEC, Ludhiana, Punjab, India, Assistant Professor, Department Of CSE & IT, Punjabi University, Patiala, Punjab, India. 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)
- [21] Payal S. Jain, Pallavi S. Panhale, Praneshwari A. Deokar, "Understanding Students Learning Experience by Data Mining of Social Media", KKWIEER, Nasik, Technical Research Organization, India.



ISSN(Online): 2320-9801  
ISSN (Print) : 2320-9798

## International Journal of Innovative Research in Computer and Communication Engineering

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 6, Issue 2, February 2018

- [22] Pranjali S. Jadhav, Dr. Shirish S. Sane, "Understanding Students learning Experience by Data Mining of social Media", KKWIEER, Nasik, Technical Research Organization, India.
- [23] Priyanka Patel and KhushaliMistry, "Classification of Student's E-Learning Experiences in Social Media via Text Mining", PIET, Vadodara, India, IOSR Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, P-ISSN: 2278-8727, Volume 17, Issue 3, PP 81-89
- [24] Ranjeeta Rana and Mrs.VaishaliKolhe, "Analysis of Students Emotion For Twitter Data using Naive Bayes and Non Linear Support Vector Machine Approach", D.Y.Patil College Of Engineering, Akurdi, SavitribaiPhule University, Pune, India. International Journal of Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 5.
- [25] Ramaswami.M and Bhaskaran.R, "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, No. 1, January 2010.
- [26] Ramesh.V, Parkavi.P, Ramar.K(2013), "Predicting student performance: A statistical and Data mining approach", International journal of computer applications, Volume 63-no. 8, PP 35-39.