Indian Journal of Science and Technology, Vol 9(48), DOI: 10.17485/ijst/2016/v9i48/103651, December 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Using Factor Classification for the Slow Learner Prediction over Various Class of Student Dataset

Swati* **and Rajinder Kaur**

Computer Science Engineering Department, Chandigarh University, Gharuan - 140413, Punjab, India;
swatisharma.cse09@gmail.com, rajinder.cse@cumail.in

Abstract

Objective: The slow learner prediction is the branch of the automatic predictive method for the students learning abilities. The high school student data has been obtained from the schools from the diverse regions in Punjab, a pivotal state of India. The selective data of almost 2400 students approximate obtained from different school in the regions has been undergone the test using the proposed model in this paper. **Method:** The analysis can be entirely performed over the student performance data. The proposed model is established upon the naïve bayes classification model for the data classification and predicts the dataset. Using the multi-factor features obtained from the input dataset. The classification algorithm has been applied individually over data grouped in the various groups of subjects. **Finding:** In this model four factor are used for predict the student performance. Two factor result are almost same but the another two factor result is different from two previous factor. The subject groups have been divided into the difficult and normal group the proposed system will be capable of performing the deep analysis over the student data obtained from high-school, the proposed model results have been show in that the deep analysis of the data tells the in-depth facts from the input data. **Improvement:** The effective accurate classification model is considered in this proposed model, when evaluated from the results described in the earlier sections.

Keyword: EDM, Factor Classification, NBC, Web Mining

1. Introduction

EDM consign into techniques, tools, and research designed for automatically mention the data by and relates to student learning activity. Stand for illustration the information is track by the various learning management systems like every student achieve all leaning article, for accessed time it or wherewith divers minutes for displayed the object in the computer screen learning of the user. The other illustration, creative guidance systems data is record the solution of problem is submitted by the whole time for learner; the submission time is collect by their, even if or not if match the result conventional result, lot of times have passed as end of submission, in which solution into the order factor was recorded by admix. The data is precision as alike rather learning environment by computer-based is describe the production big number of data is process to analysis.

Slow learner: The slow learner is no study the disability and inductive kind. If knowing about the materials those students suffer by the low rate.

The Internet provider the services to need search approach for assumption such as user's behaviors and illustrate knowledge into drop off load the traffic and Web site suite is crate to contrasting users. The trade investigator needs for study the needs of the user by the tools. Whole the carrying apparatus and approach for advice to satisfy them essential rand the numbers of counted problems are clarify at Web. Thus, Web Mining has become modish effective field.

In the prediction model, we asset out in the internet users which are looking for some user so be may be survey on the detailed data. Rules are submission of this type or facts of mining approach for search alluring user pattern by WWW rules or picture in the adjustment for realize or serve appetite to applications of web based. User

*Author for correspondence

figures or facts seizing personal either cause of WWW consumer separately with flip through in the WWW site. Web application excavation ourselves could classifying further calculate the facts and figures considered by the kind of user: Data for web server: Web server is used for client logs anthologized.

2. Review Paper

Author in[3] has to crate the techniques to WUM use to enhance the data mining algorithm that is FPT algorithms (Frequency Pattern Tree). WUM could be classifying the kind of usage data is study by the web slave, slave application or level of application data. The proposed work consternate in the WUM or correctly running on the tabs crosswise example of web application find the records of log server using site. Apriori algorithm or improved Frequent Pattern Tree algorithm are used to memory and time is bonding for comparing. Author in[2] have proposed on knowledge usage by web server logs use the WUM. Faces the major problem by any website in admin either another web application system data is raise per second, in server log files is stored the other kind or formats around the customer, according to their previous data on the website and web services for their future and maintain the structure. WUM ambition at discovering meaningful instruction and ability by the log file are certified by the user, upon constitutional different type data are use in process of mining. Using techniques from web mining, in article into procure ability by the web server log files all user exploration history is certified uses a web usage mining techniques. Authors in[4] discussed an algorithm the frequent access pattern identification for prediction model. In WM inquiry identifies is done by the web logs from customer's find formats. The general access to find formats ,tree formation designed or tested are completed using the frequency access pattern, although planned algorithm not want to create the tree. Studies of pattern is done by planned web construction, wanted just one database scan and enhance the ability for another algorithm for matching.

3. Simulation Model

The independent variables are considered for the purpose of prediction or occurrence of the event. In spite of their naive design and ostensibly made easy assumptions, naive Bayes classifiers often work much better in many complex real numbers than one might expect. Recently, careful analysis of the Bayesian classification problem has shown in the theoretical reasons for the probably unreasonable competence of naive Bayes classifiers. An advantage of the naive Bayes classifier is that it requires a miniature amount of training data to evaluate the parameters (means and variances of the variables) necessary for classification and floating point also support by the navies' bayes classification. Because independent variables are simulated, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

Algorithm 1: Naive Bayes Classifier for Slow Learner Prediction

- consider opinion="yes", opinion="no" possibility Po yes, Po no by instruction input data .
- For Test all the Input data
- For all aspect
- consider kind of aspect still upon absolute Division
- Consider possibility to opinion="Yes", opinion="No" compare classe Po(A, Yes), Po(A, No) by Training data.
- For all aspects
- consider analysis yes = Analysisyes *Po(A, Yes), analysis no = analysisno*Po(A, No);
- consider Analysis yes = Analysis yes *Po yes
- Analysis no = Analysis no *Po no;
- If (anaylsis yes > Analysis no) Then opinion ="Yes";
- otherwise opinion ="No";

The Formulae used under the Naïve Bayes classifier algorithm:

- Poyes= if all number of records are yes.
- Pono= if all number of records are no.
- Po(a,yes)= All number of yes in compare classes .
- Po(a,no)= All number of no in compare classes.

Algorithm 2: Customized Naïve bayes classification model

1. Load the student review dataset
2. Select the parameter set according to the input requirement
3. Select the naïve bayes for the classification model
4. Run the classification model
5. Initiate the iteration parameters
6. For each input record

a. For each attribute

i. The entities are divided in the separate categories according the categorical data.

ii. The probability is calculated from the training input

7. For each attributes
a. Calculate the probability and classify the data according the found irrelevance parametric setup.
b. Return the diagnosis parameters
8. Return the sentiment classification data

4. Result Analysis

4.1 Results of First Entity

The first entity (or school) data is containing the 1300 testing samples. The complete dataset has been evaluated under the proposed model for the evaluation of the student learning capabilities. The overall assessment of the students is collected using the classification of the multiple factors against the pre-evaluated dataset classification decision taken on mean averaging factor based threshold. The student classification decision has been taken in the quad-factoring based methodology under the proposed model. The quad-factoring includes the two groups of subjects, where the subjects of science and mathematics are grouped under one class and English and social science in other group. The mean and median based factors have been utilized in the classification paradigm. The other paradigms of classification require the high order averaging, which includes the factorization using the standard deviation and covariance in the fourth classification paradigm. The following figures explain the effectiveness of the proposed model in analyzing and classifying the slow learning students effectively.

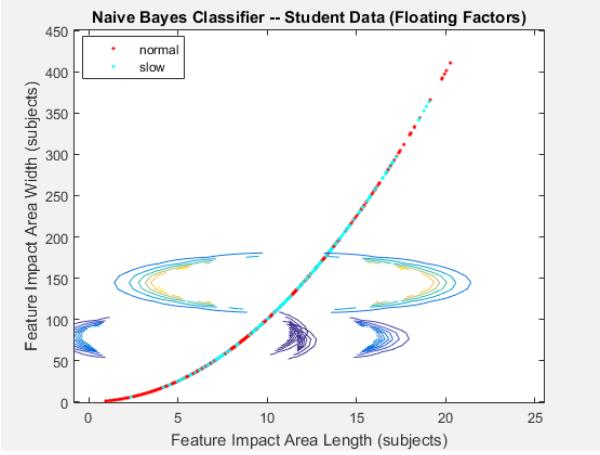

Figure 1. Naïve based classification based upon the floating factors.

Table 1. The pre-classification categorization result

Value	Count of Values	Percentage
Normal	1155	88.85
Slow	145	11.15

4.2 Results of Second Entity

The second entity (or school) data is containing the 461 testing samples. The complete dataset has been evaluated under the proposed model for the evaluation of the student learning capabilities. The overall assessment of the students is collected using the classification of the multiple factors against the pre-evaluated dataset classification decision taken on mean averaging factor based threshold. The student classification decision has been taken in the quad-factoring based methodology under the proposed model. The following figures displays the results obtained from the proposed model.

Table 2. The post-classification categorization results after the multiple variance based

Category	normal	slow
FIRST STAGE CLASSIFICATION	88.85	11.15
SUBJECT CLASS 1	40.8	10.54
SUBJECT CLASS 2	47.2	13.14
AVERAGING FACTORS	44.82	6.3
FLOATING FACTORS	11.42	3.71

Table 3. The pre-classification categorization results

Value	Count of Values	Percentage
Normal	407	88.29
Slow	54	11.71

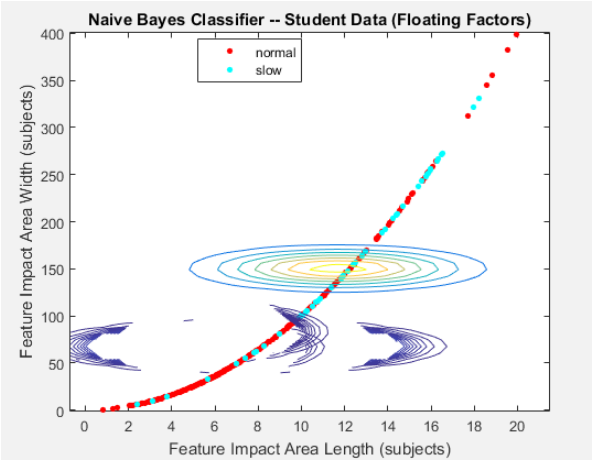

Figure 2. Naïve based classification based upon the floating factors.

4.3 Results of Third Entity

The third entity (or school) data is containing the 625 testing samples. The complete dataset has been evaluated under the proposed model for the evaluation of the student learning capabilities. The overall assessment of the students is collected using the classification of the multiple factors against the pre-evaluated dataset classification decision taken on mean averaging factor based threshold. The student classification decision has been taken in the quad-factoring based methodology under the proposed model. The following figures displays the results obtained from the proposed model.

Table 4. The post-classification categorization results after the multiple variance based result evaluation

Category	Normal	Slow
FIRST STAGE CLASSIFICATION	88.29	11.71
SUBJECT CLASS 1	41.24	9.85
SUBJECT CLASS 2	47.68	13.65
AVERAGING FACTORS	44.23	6.23
FLOATING FACTORS	11.69	3.74

Table 5. The pre-classification categorization results

Value	Count of Values	Percentage
Normal	554	88.64
Slow	71	11.36

Table 6. The post-classification categorization results after the multiple variance based result evaluation

Category	normal	Slow
FIRST STAGE CLASSIFICATION	88.64	11.36
SUBJECT CLASS 1	38.69	9.84
SUBJECT CLASS 2	47.16	13.54
AVERAGING FACTORS	43.54	7.005
FLOATING FACTORS	11.21	4.10

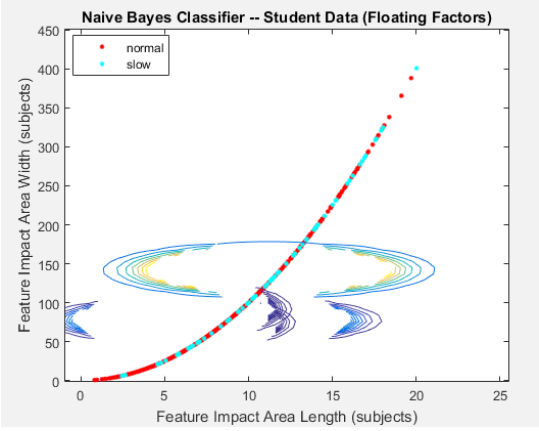

Figure 3. Naïve based classification based upon the floating factors.

5. Conclusion

The proposed model has been designed for the slow learner prediction out of the given database of the student performance. The proposed model utilizes the various factors for the analysis of the student performance data for the prediction of the slow learners. The proposed model is working in both way online and offline student data. The simulation model detects the abnormalities in the given API data the proposed model output has been designed sequential order to perform all the operations as per the system design. This model is better for classification of the student data. This is the best for the evaluation for the student results is very accurately model.

6. Reference

1. Aggarwal CC, Zhai C. A survey of text classification algorithms. Mining Text Data. US: Spirnger; 2012. p. 163–222.
2. Eltahir MA, Dafa-Alla AF. Extracting knowledge from web server logs using web usage mining. 2013 International Conference on IEEE Computing, Electrical and Electronics Engineering (ICCEEE); 2013 Aug. p. 413–7.
3. Gupta A, Arora R, Sikarwar R, Saxena N. Web usage mining using improved frequent pattern tree algorithms. 2014 International Conference on IEEE Issues and Challenges in Intelligent Computing Techniques (ICICT); 2014 Feb. p. 573–8.
4. Sharma MM, Bala A. An approach for frequent access pattern identification in web usage mining. ICACCI, 2014 International Conference on IEEE Advances in Computing, Communications and Informatics; 2014 Sep. p. 730–5.
5. Ting SL, Ip WH, Tsang AH. Is Naive Bayes a good classifier for document classification? International Journal of Software Engineering and Its Applications. 2011; 5(3):37–46.
6. Baker RS. Educational data mining: An advance for intelligent systems in education. IEEE Intelligent Systems. 2014; 29(3):78–82.
7. Sajana T, Rani CMS, Narayana KV. A survey on clustering techniques for big data mining. Indian Journal of Science and Technology. 2016 Jan; 9(3):1–12.
8. Vedanayaki M. A study of data mining and social network analysis. Indian Journal of Science and Technology. 2014 Nov; 7(S7):1–3.

9. Sudha M, Kumaravel A. Performance comparison based on attribute selection tools for data mining. Indian Journal of Science and Technology. 2014 Nov; 7(S7):1–5.
10. Midhunchakkaravarthy J, Brunda SS. An enhanced web mining approach for product usability evaluation in feature fatigue analysis using LDA model and association rule mining with fruit fly algorithm. Indian Journal of Science and Technology. 2016 Feb; 9(8):1–10.
11. Asha KN, Rajkumar R. Survey on web mining techniques and challenges of e-commerce in online social networks. Indian Journal of Science and Technology. 2016 Apr; 9(13):1–5.