iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** VII **Month of publication:** July 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

Review of Different Algorithms to Pridict Academic Attrition

Vishal Mittal[1], Anuradha[2]
[1,2,]*Department of Computer Science & Engineering, Shree Siddivinayak Group of Institutions, Bilaspur, Yamuna Nagar, Haryana*

Abstract:** Quality of education system is very important for a country growth. Today education sector is facing challenges, the major challenges of higher education being decrease in students success rate and their leaving a course without completion. An early prediction of student's failure can avoid poor performance, which will help to enhance their performance. It can help not only the current students but also the future students to predict thier performance. Data mining provides powerful techniques to analysis student performance. For this purpose, In this dissertation various educational data mining techniques have been used such as Naive Bayes, Decision Tree, K-Nearest Neighbour, Random Forest, Rpart , C5.0 to build a model for academic attrition based on students social integration, academic integration and various emotional skills considered. In order to future through data mining techniques data was collected from mullana university. Data from the admission process are complemented with the academic information that is gathered for each academic period; however, the causes of low academic performance occur on day-to-day basis and waiting until the academic period ends could be crucial. This leads to think that new, and possibly, non – traditional ways, for collecting information close to real time are needed. In this dissertation new attributes are identified which represent real time student academic attrition. The implementation of different data mining techniques is done on R language develop at the university of Auckland, New Zealand. It is an open source language. It is an interactive language used for easy input,output and large data manipulation and used for various statistical analysis and modelling. Many classification and regression algorithms are used, which are attribute dependent. Some are used on categorical, nominal data others on numerical data. The experimental results are validated against test data and interesting co-relations are observed. The comparison of their accuracy is done to find the most accurate predictions. Graphs are also used for illustrative comparison, along with numerical values.

Keywords: R language; data mining; attributes

I. INTRODUCTION

Education means obtaining knowledge. A person who has knowledge of his surrounding can survive happily in the society. To get acquired with it, people join educational institutes, were time and money both being extremely precious thing themselves, must be spent efficiently. Higher education has gained important manifolds in the past few decades. The higher education institute is forced to revise its scope and objects because of the private participation. Universities today, similar to business organizations, are operating in a very dynamic and strongly competitive environment. The education globalization leads to more and better opportunities for students to receive high quality education at institutes all over the world. [1].For higher education institution whose goal is to contribute to the improvement of quality of higher education, the success of creation of human capital is the subject of a continuous analysis. Therefore, the prediction of student's success is crucial for higher education institutions, because the quality of teaching process is the ability to meet student's needs. In the sense important data and information are gathered on a regular basis, and they are considered at the appropriate authorities, and standards in order to maintain the quality are set. The quality of higher education institutions implies providing the services, which most likely meet the needs of students, academic staff, and other participants in the education system, the participants in the educational process, by fulfilling their obligations through appropriate activities, create an enormous amount of data which needs to be collected and then integrated and utilized. By converting this data into knowledge, the gratification of all participates is attained: students, professors, administration, supporting administration, and social community.[2]

All participants in the educational process could benefit by applying data mining on the data from the higher education system (figure 1.1). Since data mining represents the computational data process from different perspectives, with the goal of extracting implicit and interesting samples (written and frank, 2000), trends and information from the data, it can greatly help every participant in the educational process in order to improve the understanding of the teaching process, and it centres on discovering, detecting and explaining educational phenomenon's (EI- Halees, 2008) [3]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

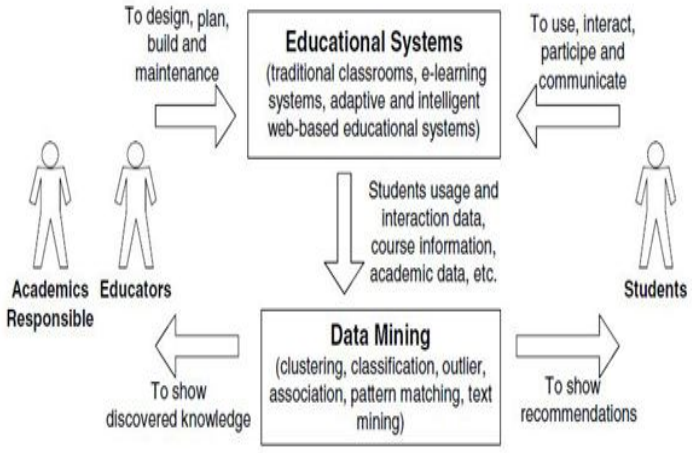


Fig 1.1 The cycle of applying data mining in education system[7]

So with data mining techniques, the cycle is built in educational system which consists of forming hypotheses, testing and training.Thus, application of data mining in educational systems can be directed to support the specific needs of each of the participants in the educational process. The student is required to recommended additional activities, teaching materials and tasks that would favour and improve his/her learning.

II. LITERATURE SURVEY

The objective of the Literature review was to explore the various approaches for student performance prediction and Academic Attrition using various techniques of data mining, exercised by various researchers. It goes on to list the most common tasks in educational environment that have been resolved through data mining techniques. The main goal of this Literature review was to discover which techniques were the most efficient one and therefore, possible solutions could be find out to overcome problems associated with them.

A. Pamela Chaudhury et al.[1]

Described that in the age of information and communication technology, technology is being used in every domain. Education is the integral part of our society consisting of the teaching, learning and evaluation process. Information and communication technologies are being used for the purpose of e–learning, measuring student's learning, course design, student performance evaluation. Using machine learning techniques performance of the students has been studied and useful results have been derived. Predicting the performance of a student accurately in the upcoming exam is of extreme significance. Every machine learning tool heavily depends upon the input data .Studying and implementing the elaborate feature set for students has improved the accuracy of the prediction system. Further use of pre-processing techniques along with classification algorithms has significantly improved the results of the prediction system.

B. Tripti Mishra et al. [2]

Introduced that A country's growth is strongly measured by quality of its education system. Education sector, across the globe has witnessed sea change in its functioning. Today it is recognized as an industry and like any other industry it is facing challenges, the major challenges of higher education being decrease in students' success rate and their leaving a course without completion. An early prediction of students' failure may help the management provide timely counselling as well coaching to increase success rate and student retention. We use different classification techniques to build performance prediction model based on students' social integration, academic integration, and various emotional skills which have not been considered so far. Two algorithms J48 (Implementation of C4.5) and Random Tree have been applied to the records of MCA students of colleges affiliated to Guru Gobind Singh Indraprastha University to predict third semester performance. Random Tree is found to be more accurate in predicting performance than J48 algorithm.

C. Priyanka Saini et al. [3]

Introduced that In recent years, Indian higher educational institutes grow rapidly. There is more competition between institutes for attracting students to get enrolment in their institutes. The admission process is conducted every year at the institute and it results in the recording of large amounts of data. But, in most of the cases this data is not properly utilized (or analyzed) and results in wastage of what would otherwise be one of the most precious assets of the institutes. By applying the various data mining techniques on this data one can get valuable information and predictions can be done for the betterment of the admission process. This study presents data mining techniques for the enrolment process in MCA stream. These methods will help to improve the overall performance of the admission process at higher educational institutes.

D. Kamaljit Kaur et al. [4]

Presented that Recently the University Grants Commission of In- dia has introduced a multistage examination system in higher education institutes in the country. The new system, called the Credit Based Continuous Evaluation and Grading System (CBCEGS), assesses a student on the basis of her continuous evaluation during the semester, combined with her performance in the end semester examination. This multistage examination pattern provides an opportunity to students to improve their performance. If a student cannot perform well in tests during the semester, she can improve her performance in the end semester test. But it does not seem so easy. In certain courses, due to their difficulty level such as mathematics, a student may not be able to improve her knowledge at the last moment despite hard work. Though, it may be possible in case of courses that are comparatively easy such as System Analysis and Design. This paper analyzes and predicts student's performance using data mining techniques for two data sets of 1000 students each one for Mathematics, and the other for System Analysis, and Design. This study can help the education community to understand learning behaviour of students as far as courses of varying difficulty are concerned. It is observed that Classification and Regression Tree (CART) supplemented by AdaBoost is the best classifiers for the prediction of students' grades for both subjects. J48 supplemented by AdaBoost performs excellent for System Analysis and Design but performs worst for mathematics and M5P generates best results for early prediction of students' marks in the major test.

E. Amjad Abu Saa et al. [5]

Provided an overview that It is important to study and analyse educational data especially students' performance. Educational Data Mining (EDM) is the field of study concerned with mining educational data to find out interesting patterns and knowledge in educational organizations. This study is equally concerned with this subject, specifically, the students' performance. This study explores multiple factors theoretically assumed to affect students' performance in higher education, and finds a qualitative model which best classifies and predicts the students' performance based on related personal and social factors.

F. Shahiri et al. [6]

Have also provided an overview on several techniques of data mining that were applied to predict and analyze performance of students, concentrating on the identification of most valuable attributes in a student's data by employing the prediction algorithm**.** They provide a systematic literature review to improve the student's achievements by using the techniques of data mining. The various analytical methods used cumulative grade point average (CGPA) as their data sets, thus helping the system of education to monitor the performance in a very systematic way.

G. Osmanbegović et al. [7]

Applied three supervised data mining algorithms to assess the data of first year students to predict favourable outcome in a course and evaluating the performance based on certain factors like convenience, accuracy and approach of learning**.** A very high emphasis is given on some socio-demographic factors, high school results obtained, attitude towards study and marks in entrance examinations. The whole data was collected from University of Tuzla, academic year 2010-2011. The authors believe that exams play a very important role to determine the future of the students, in addition to the internal assessments. They used WEKA for their study and implemented it in java and also conducted four tests to assess the input variables: Info Gain test, Chi-square test, Gain Ratio test and One R-test.

H. Prof. R. A. Gangurde et al. [8]

Presented that Data mining is a logical process which finds useful patterns from large amount of data. It is the process of extracting previously unknown, comprehensible and actionable information from large databases and using it to make crucial business decisions. Data mining is the computer-assisted process that digs and analyzes enormous sets of data and then extracts the knowledge out of it. The various techniques of data mining are used to extract the useful piece of knowledge from a database / data warehouse which is growing continuously. This extraction of knowledge is useful in research as well as in organization. In this paper authors have reviewed the literature of data mining techniques such as Classification, Clustering, Association Rules and Prediction.

I. Yiming Ma, Bing Liu et al. [9]

Described that education domain offers a fertile ground for many interesting and challenging data mining applications. These applications can help both educators and students, and improve the quality of education. In this paper, we present a real-life application for the Gifted Education Programme (GEP) of the Ministry of Education (MOE) in Singapore. The application involves many data mining tasks. This paper focuses only on one task, namely, selecting students for remedial classes. Traditionally, a cut-off mark for each subject is used to select the weak students. That is, those students whose scores in a subject fall below the cut-off mark for the subject are advised to take further classes in the subject. In this paper, we show that this traditional method requires too many students to take part in the remedial classes. This not only increases the teaching load of the teachers, but also gives unnecessary burdens to students, which is particularly undesirable in our case because the GEP students are generally taking more subjects than non-GEP students, and the GEP students are encouraged to have more time to explore advanced topics. With the help of data mining, we are able to select the targeted students much more precisely.

J. Ramesh et al. [10]

Proposed A model has also been developed based on some selected input attributes assembled through questionnaire method conducted a survey cum experimental methodology to generate database for the students for predicting the performance. The three main objectives were to identify the essential predictive variables on higher secondary students, know the best classification algorithm and to predict the grade at higher examinations. The study shows that parent's occupation plays a major role and not the type of school in predicting the grades. The data for the study was collected from schools and internet and the authors found out that multilayer perceptron algorithm is the best one for grade prediction. This algorithm is more efficient showing the accuracy of 72%.

III. PROCEDURE FOR GRADE CALCULATION

The data mining classification algorithms are different in many aspects such as: the learning rate, performance, speed, correctness, robustness, accuracy, etc. In this research, we examined thoroughly the impact of five algorithms for performance prediction: Decision tree, C5.0, Naive bayes, Random forest, KNN algorithm. The five classification techniques are employed to reveal the most appropriate way to measure student's performance.

A. Collection of all the grades attained by each student in same sequence of subjects, as shown in Fig 5.1. The grades are collected as GRADES(i), where 'i' is consecutive exams conducted and $1 \leq i \leq 4$. Grades(i)=consecutive exam, $1 \leq i \leq 4$
E.g.: Grade(2)= {B, A, B, B, C, A} is the sequence of subject grades of the student in second consecutive Exam having enrolment no. 130507009 and studying six subjects as highlighted in Fig. 3.1.

ID No	GRADES(1)	GRADES(2)	GRADES(3)	GRADES(4)
130507001	A,D,B,D,D,C	D,B,B,D,C,D	D,D,C,D,D,D	B,C,D,D,C,C
130507002	C,B,B,B,C,B	C,C,C,C,C,C	B,B,B,B,B,B	A,B,B,B,A,A
130507003	D,C,C,B,B,C	D,D,E,D,D,C	B,C,D,D,C,C	B,B,B,C,B,B
130507004	C,D,B,C,C,B	D,C,D,B,C,B	D,B,C,B,B,C	A,B,C,B,C,B
130507005	B,B,A,A,B,A	A,A,B,B,A,A	A,A,B,B,A,A	B,A,B,A,B,A
130507006	C,C,E,C,D,D	C,D,D,D,F,D	D,C,E,D,D,E	C,C,C,C,C,C
130507007	B,B,B,B,B,B	B,B,B,C,B,B	A,B,C,B,B,B	D,B,D,B,B,B
130507008	C,D,D,C,D,B	D,D,F,D,D,D	C,D,D,C,D,D	C,C,E,C,D,C
130507009	A,A,B,B,A,A	B,A,B,B,C,A	A,A,B,A,A,A	C,C,E,C,D,C
130507010	B,CD,D,C,C	D,D,D,F,D,C	C,D,C,C,C,D	C,D,D,C,C,C

Fig. 3.1. Collection of grades for four Exams

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 5 Issue VII, July 2017- Available at www.ijraset.com

The grades and their corresponding performance criteria is shown in Table 3.1. This performance criteria is used for prediction in the final outcome i.e. OVERALLGRADE(F), where 'F' stands for final.

B. Prepare the logic table (n^{th} level logic predicate) on the basis of GRADE in Table 3.1.

Level-wise logic order:

L_0 : A – A → A

D – D → D

B – A → A

B – B → B

A – B → A

L_0 : A – C → B

B – D → C

C – E → D

L_1 : A – D → B

B – E → C

D – A → C

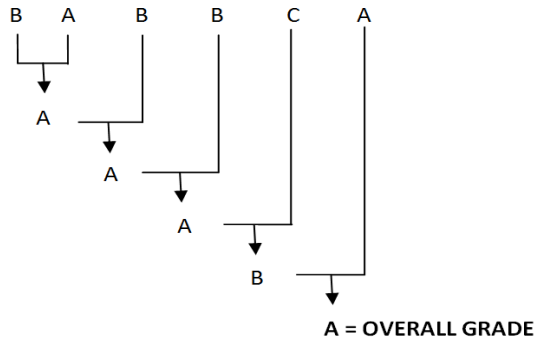
E – B → D

Table 3.1: Performance Criteria

Grade	Performance	Marks
A	Excellent	9-10
B	Good	8-7
C	Average	6-5
D	Poor	4-3
E	Fail	2-1

L3: A-EC

E.g.: We have the sequence of grades {B, A, B, B, C, A} for the student having enrolment no. 130507009. While applying the step 2, take two consequent grades . together and then compute the output in the way as shown below using Table 3.1. Like, B and A gives output A, then take this output as input for the next step. Now A and B gives output A. These steps are done using the above level-wise logic order until we reach to our final grade.



This shows that the performance of the student is Excellent, on the basis of Table 3.1.

C. Therefore, OVERALLGRADE(i), $1 \leq i \leq 4$ is computed for each semester(in the same manner as used above), where *i* is the semester. E.g. OVERALLGRADE(2) is the overall grade computed for second Exam.

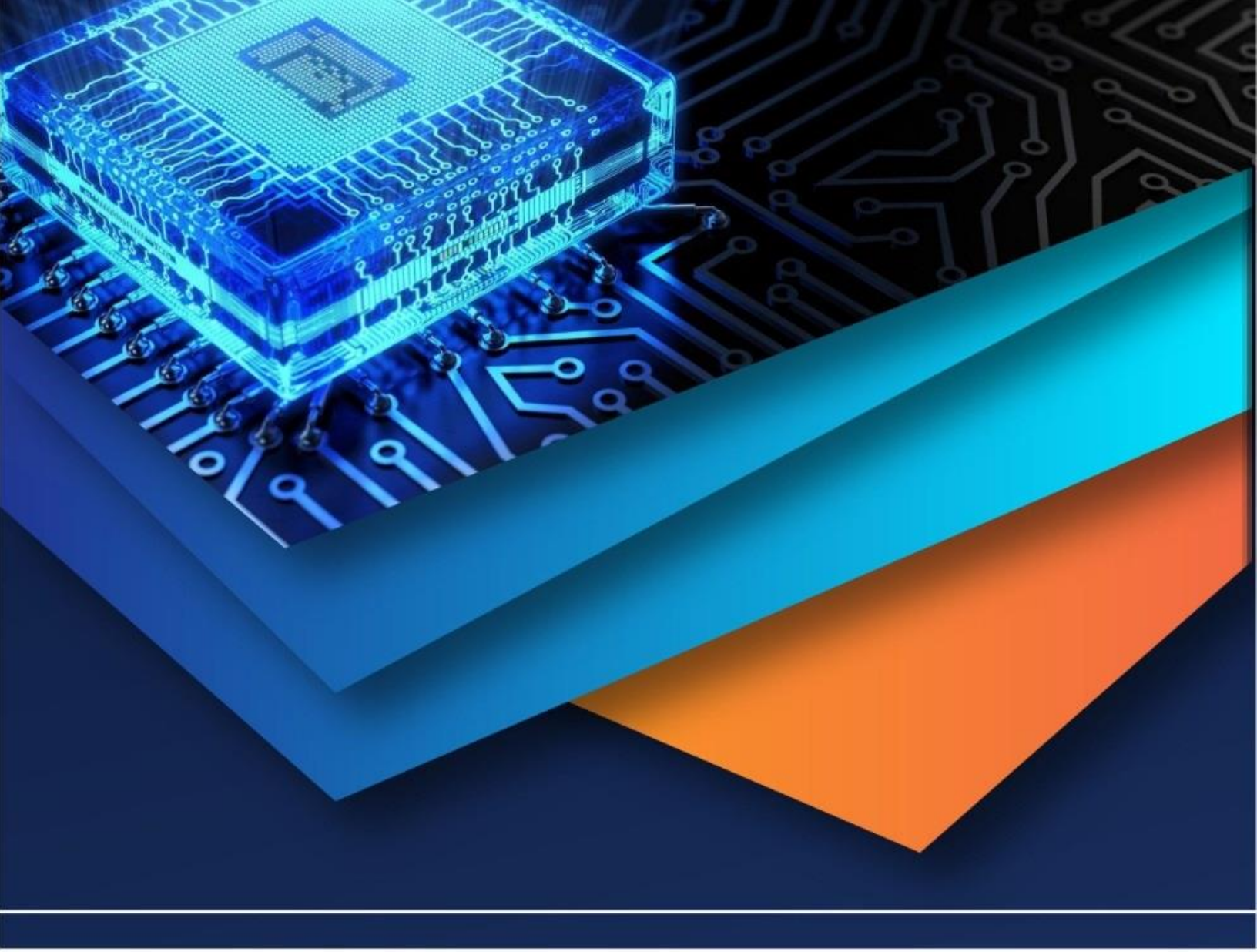
ID No	GRADES(1)	OVERALL Grade(1)	GRADES(2)	OVERALL Grade(2)	GRADES(3)	OVERALL Grade(3)	GRADES(4)	OVERALL Grade(4)
130507001	A,D,B,D,D,C	C	D,B,B,D,C,D	C	D,D,C,D,D,D	D	B,C,D,D,C,C	C
130507002	C,B,B,B,C,B	B	C,C,C,C,C,C	C	B,B,B,B,B,B	B	A,B,B,B,A,A	A
130507003	D,C,C,B,B,C	B	D,D,E,D,D,C	C	B,C,D,D,C,C	C	B,B,B,C,B,B	B
130507004	C,D,B,C,C,B	B	D,C,D,B,C,B	B	D,B,C,B,B,C	B	A,B,C,B,C,B	B
130507005	B,B,A,A,B,A	A	A,A,B,B,A,A	A	A,A,B,B,A,A	A	B,A,B,A,B,A	A
130507006	C,C,E,C,D,D	C	C,D,D,D,F,D	D	D,C,E,D,D,E	D	C,C,C,C,C,C	C
130507007	B,B,B,B,B,B	B	B,B,B,C,B,B	B	A,B,C,B,B,B	B	D,B,D,B,B,B	B
130507008	C,D,D,C,D,B	B	D,D,F,D,D,D	D	C,D,D,C,D,D	D	C,C,E,C,D,C	C
130507009	A,A,B,B,A,A	A	B,A,B,B,C,A	A	A,A,B,A,A,A	A	C,C,E,C,D,C	B
130507010	B,C,D,D,C,C	C	D,D,DF,D,C	C	C,D,C,C,C,D	C	C,D,D,C,C,C	C

Fig 3.2 Computation of Overall grade for each exams

D. Similarly, OVERALLGRADE(F) for each of the student is computed as the final performance result

REFERENCES

[1] C. Romero, S. Ventura, “Educational data mining: a survey from 1995 to 2005”, Expert Systems with Applications, vol. 33, no.1, pp.135-146, 2007.
[2] D. P. Nithya, B. Umamaheswari, A. Umadevi, "A survey on educational data mining in field of education," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 5, no. 1, pp. 69–78, Jan. 2016.
[3] E. Osmanbegović, M. Suljić, "Data mining approach for predicting student performance", Economic Review – Journal of Economics and Business, vol. 1, no. 1, pp. 3-12, 2012.
[4] R. Srivastava, M. Gendy, M. Narayana, Y. Arun, J. Singh, University of the future — “A thousand year old industry on the cusp of profound change”, Melbourne, Australia: Ernst & Young (Retrieved from http://www.ey.com/Publication/vwLUAssets/University of the future/$FILE/University of the future 2.12.pdf), 2012.
[5] Y. Ma, B. Liu, C. Wong, P. Yu, S. Lee, “Targeting the right students using data mining”, In: Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data mining, pp. 457–464, 2000.
[6] D. Suthers, K. Verbert, E. Duval, X. Ochoa, “Clow, MOOCs and the funnel of participation”, In: International Conference on Learning Analytics and Knowledge, pp. 185–189. ACM New York, 2013.
[7] Suthers, K. Verbert, E D. Silva, M. Vieira, “Using data warehouse and data mining resources for ongoing assessment in distance learning”, In: IEEE International Conference on Advanced Learning Technologies, pp. 40–45, 2002.
[8] J. Anderson, A. Corbett, K. Koedinger, “Cognitive tutors”, J. Learn. Sci., vol.4, no.2, pp.67–207, 1995.
[9] P. Brusilovsky, C. Peylo, “Adaptive and intelligent web-based educational systems”, Int. J. Artif. Intell. Educ. 13, vol.2, no.4, pp. 159–172, 2003.
[10] C. Romero, S. Ventura, E. Salcines, “Data mining in course management systems: moodle case study and tutorial”, Comput. Educ., vol.51, no.1, pp. 368–384, 2008.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY