



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: II Month of publication: February

DOI: <http://doi.org/10.22214/ijraset.2019.2172>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

A Study on Educational Data Mining

Ms. Falguni Suthar¹, Ms. Hiralben Patel², Dr. Bhavesh Patel³

^{1, 2, 3}Acharya Motibhai Patel Institute of Computer Studies, Ganpat University

Abstract: Educational data mining is an emerging field that focuses on analysing educational data to develop models to improve learning experiences and improve institutional effectiveness. It provides inherent knowledge about the delivery of education, which is used to improve the quality of teaching and learning. Effective planning can provide personalized education. The mining of educational data develops new methods to discover the knowledge of the educational database and is used for decision making in the education system. This document presents a study on various components of educational data mining along with its objectives. The objective of this document is to present a brief general description of EDM and observe the development in the field of EDM.

Keywords: EDM, learning, teaching, education, mining.

I. INTRODUCTION

Educational data mining deals with the development, research and application of computerized methods to detect patterns in large collections of educational data that would otherwise be difficult or impossible to analyse due to the enormous volume of data within which they exist [5] In recent years, there is an exponential growth in the education sector that leads to the growth of educational data, so that the analysis of educational data becomes an important aspect.

EDM analyses the data generated by any type of information system that supports learning or education (in schools, colleges, universities and other academic or professional learning institutions that offer traditional and modern forms and methods of teaching, as well as informal learning). This data is not limited to the interactions of individual students with an educational system (for example, navigation behaviour, contributions in tests and interactive exercises), but may also include data from collaborating students (for example, text chat), data administrative (for example, school, school) district, teacher), demographic data (for example, gender, age, school grades), student's affectivity (for example, motivation, emotional states), etc. These data have typical characteristics such as multiple levels of hierarchy (subject, assignment, question levels), context (a particular student in a particular class that finds a particular question at a particular time on a particular date), grainy (record of data in different resolutions to facilitate different analysis, for example, data recording every 20 seconds), and longitudinal (much of the data recorded in many sessions over a long period of time, for example, covering semester courses and of one year) [5].

The EDM can be drawn as the combination of three main areas (Fig. 1): informatics, education and statistics. The intersection of these three areas also forms other sub-areas closely related to the EDM, such as computer-based education, machine and machine learning, and learning analysis (LA)[5].

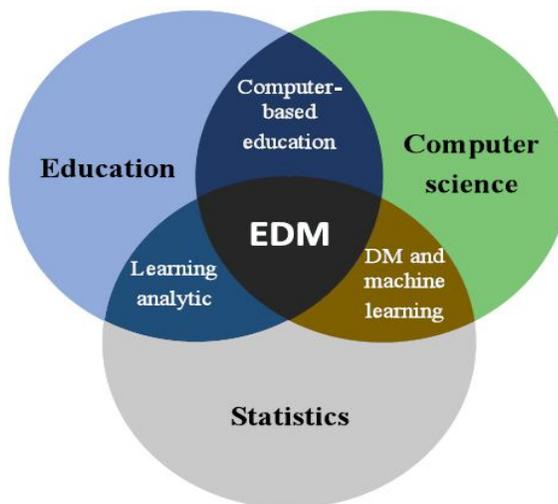


Fig.1 main area involved in Educational data mining.

Source: Cristobal Romero & Ventura, (2013)

The main objective of EDM is to improve the educational system and the EDM objectives are student / student behavior models, prediction of performance, increase of reflections and knowledge, prediction of dropouts and retention, improvement of evaluations and feedback services, recommendation of resources [6] The education system involves different groups of users or participants. Educational data Decision making in the field of academic planning involves an extensive analysis of large volumes of educational data. The data is generated from heterogeneous sources as diverse and distributed, structured and unstructured data. This data is generated mainly from offline or online sources: Data offline. Offline data is generated from a traditional and modern classroom, interactive teaching / learning environments, information about learners, student attendance, emotional data, course information, data collected from the academic section of an institution, etc. Online data is generated from separate geographic participation in education, distance education, web-based education and computer-assisted collaborative learning used on social networking sites and online group forums. For example: web logs, email, spread sheets and phone conversations with scripts, medical records, legal information, corporate contracts, text data, publication databases, etc. [1].

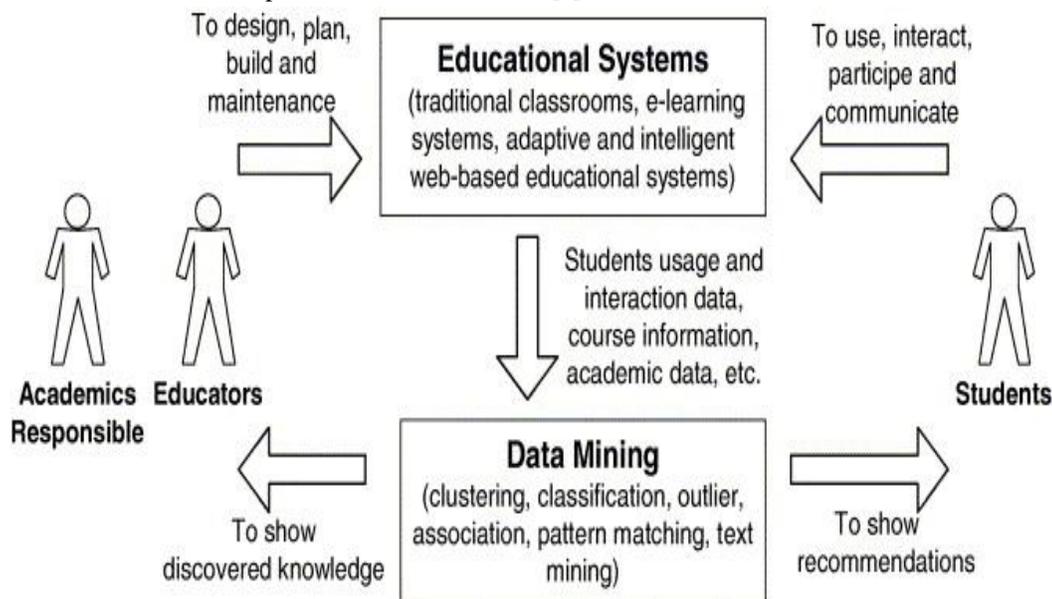


Fig. 2 Need of Educational Data Mining [12]

In Fig. 2 (EDM) we represent the need for educational data mining. Academics and educators worked in the education system to improve student performance. This diagram shows that

Educators want to design the education system, and then plan to build that system and, more importantly, maintain that educational system. Educational systems include traditional classrooms and some innovative learning methods such as the e-learning system, the intelligent and adaptable web-based educational system, etc.

The data set can be extracted from the students, since the students connect directly with the educational system. Now the data is provided as input to the data extraction process and, consequently, offers recommendations to the students and extracts new knowledge from the educators by using various data extraction techniques such as grouping, classification, comparison of patterns, etc. [2].

II. GOALS OF GOALS OF EDUCATIONAL DATA MINING IN HIGHER EDUCATION

- 1) Predicting student's future learning behaviour: with the use of student models, this goal can be achieved by creating models of students that incorporate student characteristics, including detailed information such as their knowledge, behaviour and motivation to learn.
- 2) Discover or improve domain models: through the various methods and applications of EDM, it is possible to discover new and improve existing models.
- 3) Study of the effects of educational support: it can be achieved through learning systems.
- 4) Advance in scientific knowledge about learning and apprentices: building and incorporating student models, the field of EDM research and the technology and software used.

III.OBJECTIVES OF EDUCATIONAL DATA MINING (EDM)

EDM aims to improve various aspects of the education system. The objectives of EDM depend on the point of view of the end users (apprentice, educator, administrator and researcher) and help solve their problems:

- 1) *Student model*: the user model in the educational domain incorporates detailed information such as student characteristics or states such as knowledge, skills, motivation, satisfaction, metacognition, attitudes, experiences and learning progress, or certain types of problems that they negatively affect learning outcomes. The common goal here is to create or improve a student model from usage information.
- 2) *Predictive modeling*: Predict student performance and learning outcomes. The goal is to predict a student's final grades or other types of learning outcomes (such as retention in a college program or future ability to learn) based on data from the course activities.
- 3) *Generation of recommendations*: the objective is to recommend to the students the content (or tasks or links) that is most appropriate for them at the present time.
- 4) *Analyze the learner's behavior*: This takes several forms: Apply data mining educational techniques to analyze the behavior of the learner.
- 5) *Maintenance and improvement of the courses*: the objective here is to determine how to improve the courses (contents, activities, links, etc.), using information (in particular) on the use and learning of the students. Discover or improve the models that characterize the topic to be learned (for example, mathematics, science, etc.), identify fruitful pedagogical sequences and suggest how these sequences could be adapted to the needs of the student.
- 6) *Students*: to support a student's reflections on the situation, to provide comments or recommendations adaptable to the students, to respond to the needs of the student, to improve the performance of learning, etc.
- 7) *Educators*: to understand the learning processes of their students and reflect on their own teaching methods, to improve the performance of teaching, to understand social, cognitive and behavioral aspects, etc.
- 8) *Administrators*: evaluate the best way to organize institutional resources (human and material) and their educational system [1].

IV.PHASES OF EDUCATIONAL DATA MINING

Educational data mining deals with the translation of new hidden information from raw data collected from education systems. EDM generally consists of the following phases:

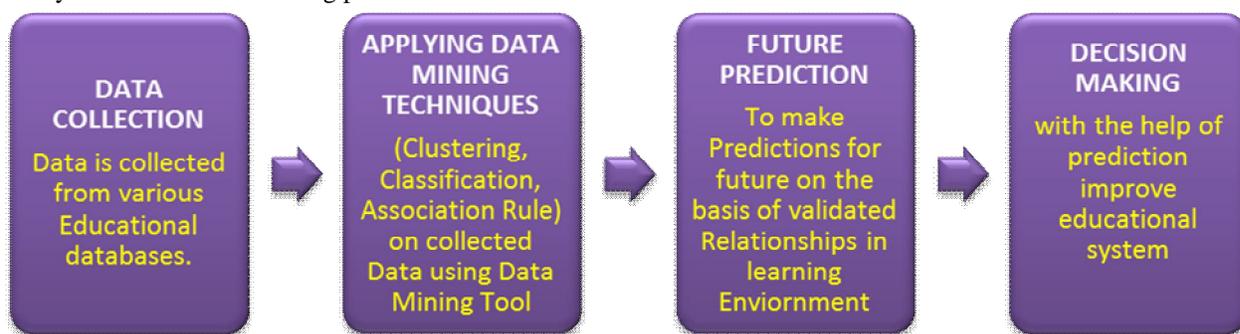


Fig. 3 Phases of EDM

The data is collected and must be extracted from different resources of the education system, i.e. the course management system (different institutes), the e-learning environment, and the web-based data (i.e. YouTube, Twitter) that they are relevant to the activities of the students during the learning process (i.e., their academic grades, student publications on social networks, etc.). Fig.3 shows the Phases of Educational Data Mining which are:

- 1) The first phase of the extraction of educational data is to find the relationships between the data of the educational environment using techniques of data extraction that is, classification, grouping, regression, etc.
- 2) The second phase of the extraction of educational data is the validation of the relationships discovered between the data to avoid uncertainty.
- 3) The third phase is to make predictions for the future based on validated relationships in the learning environment.
- 4) The fourth phase is to support the decision-making process with the help of predictions [2].

V. COMPONENTS OF EDUCATIONAL DATA MINING

Educational data mining touches and affects many aspects of the education industry. The main components of EDM are: education stakeholders, various tools and techniques of data extraction, educational data, environment and educational task and how they meet educational objectives.



Fig. 4 components of EDM [9]

- 1) *Stakeholders*: Taking into account all aspects of education, that is, primary to higher education, stakeholders in education can be classified mainly into the following:
- 2) *Learners/Students*: the most important and impacted component are the students. As students directly participate in the learning process, they fall into the main stakeholder group. The EDM can help them with personalized education based on several recommendations and can increase the interest of education for students towards learning. Different learning tasks can be formulated in different groups of students according to their needs.
- 3) *Faculties*: Educators, teachers and instructors benefit as they can determine which student needs additional support. Predicting student performance becomes easy. Another impact is that it helps in classifying students in groups. It can also provide insight into patterns in which students can learn: regular and irregular. Teachers can analyze the data and determine the most common mistakes. Beyond the academic, the analysis of the student's learning and behaviour can also be done to detect if they require any additional support during the learning process. The teachers are also the main stakeholders.
- 4) *Parents*: Parents are part of the secondary group. They are responsible for helping their children enroll in the courses most suitable for them.
- 5) *Course Researchers and Educational Developers*: they are the people who design and modify the course. They are responsible for the growth of education. The developers enter the group of secondary stakeholders.
- 6) *Administrators*: They can also be called as hybrid users. EDM is useful for the effective use of resources; It can help determine which offers can capture more students in various programs and courses. They are responsible for several administrative decisions, such as the development of infrastructure and the employment of expert teachers.

VI. EDUCATIONAL DATA MINING METHODS

There are so many promoted methods of educational data mining but all kind of methods lie in one of following specified categories:[10]

- 1) *Prediction*: Ryan S. J. d. Baker has given a detailed explanation of the prediction in his article. He mentioned that "in prediction, the goal is to develop a model that can infer a single aspect of the data from some combination of other aspects of the data. If we study the prediction exhaustively, we obtain three types of prediction: classification, regression and density estimation. In any prediction category, the input variables will be categorical or continuous. In the case of classification, categorical or binary variables are used, but continuous input variables are used in the regression. The estimation of the density can be done with the help of several functions of the Kernel.
- 2) *Clustering*: In clustering technique, the data set is divided in various groups, known as clusters. When data set is already specified, then the clustering is more useful. As per clustering phenomenon, the data point of one cluster and should be more similar to other data points of same cluster and more dissimilar to data points of another cluster. There are two ways of initiation of clustering algorithm. Firstly, start the clustering algorithm with no prior assumption and second is to start clustering algorithm with a prior postulate.

- 3) *Relationship Mining*: Relationship mining generally refers to creating new relationships between variables. It can be done in a large data set, having a no of variables. Relationship mining is an attempt to discover the variable that is most closely associated with the specified variable. There are four types of relationship mining: association rule mining, correlation mining and sequential pattern mining, and causal data mining. Association data mining is based on the if-then rule, which is if a particular set of variable values appears, and then it usually has a specific value. In correlation mining, linear correlations are discovered between variables. The goal of sequential pattern mining is to extract temporal relationships between variables.
- 4) *Discovery with Models*: includes the design of models based on some concepts such as prediction, grouping and knowledge engineering, etc. These predictions of newly created models are used to discover a new predicted variable.

VII. TOOLS USED IN EDUCATIONAL DATA MINING

The most commercial data mining tools which provide multiple data mining functions and multiple knowledge discovery techniques and the environment on which they run are listed in Table 1.

TABLE I
TOOL USED IN EDM

Name of Tool and Developer	Source (Commercial)	Function/Features	Techniques/Tools	Environments
Intelligent Miner (IBM)	Commercial	Provides tight integration with IBB's DB2 relational database system, scalability of mining algorithm	Association mining, classification, regression, predictive modelling, deviation detection, clustering, sequential pattern	Windows, solaris, Linux
MSSQL Server 2005 (Microsoft An)	Commercial	Provides DM functions both in realtional db system and Data Warehouse (DWH) system environment	Integrates the algorithms developed by third party vendors and application users.	Windows, Linux
sPSS clementine (IBM)	Commercial	Provides an integrated data mining development environment for end users and developers.	Association mining, clustering, classification, prediction and visualization tools	Windows, Solaris, Linux
Enterprise Miner (SAS Institute)	Commercial	Provides variety of statistical analysis tools	Association mining, classification, regression, time series analysis, statistical analysis, clustering	Windows, Solaris, Linux, Insightful Miner
Oracle data mining (oracle Corporateion)	Commercial	Provides an embedded DWH infrastructure for multidimensional data analysis	Association mining, classification, prediction, regression, clustering, sequence similarity search and analysis	Windows, Mac, Linux
DBMiner (DBMiner technology Inc.)	Commercial	Provides multiple mining algorithms, data-cube based on-line analytical mining, frequent pattern mining functions and integrated visual classification methods	Discovery – driven OLAP analysis, association, classification and clustering	Windows, Linux

VIII. DEVELOPMENT IN THE FIELD OF EDM

In “Educational data mining: A survey from 1995 to 2005” C. Romera and S. Ventura mentioned the development in the field of EDM till 2005. In “Understanding Educational Data Mining (EDM)” Mr.Suhas G. Kulkarni, Mr.Ganesh C. Rampure, Mr.Bhagwat Yadav has mentioned the development in the field of EDM from 2005 to 2012. Here Table 2 gives a brief idea about the major developments took place after 2012.

TABLE 2
DEVELOPMENT IN THE FIELD OF EDM

Author(year)	Work Carried
S. Borkar and K. Rajeswari (2013)	By applying Association rules, Predicting students' performance - Student's performance level can be Improved in university result by identifying students who are poor unit Test, Attendance, assignment and graduation and giving them additional guidance to improve the university result.
R. Jindal and M. Borah (2013)	By applying Clustering & Classification methods, Student performance Prediction - The students to perform poor in academic activities and it even leads to course drop outs.
M. Chakurkar and P. Adiga (2014)	By applying Usage Mining & Content Mining, Personalized recommendation based on users browsing history - Better performance improvement
P. Ratnapala, R. G. Ragel and S. Deegalla (2014)	By applying Clustering method, Students interaction with an e-learning system through instructor -led non graded and graded courses - The majority of the student populations are not self-motivated to do self-learning. (Self-Directed Learning)
T. Z. and A. M.Mahmoud (2014)	By applying Clustering method, The clustering of elementary school slow learner students behaviour for the discovery of optimal learning patterns - Enhance student learning capacities
K. Rangra and K. Bansal (2014)	By applying Classification method, Predicting students' Academic performance - Take early actions to help and assist the poor and average category students to improve their results.
A. Bogarín, C. Romero and R. Cerezo (2015)	By applying Clustering method, Discover students' navigation paths or trails in Moodle - obtain more specific and accurate trails.
A. Abu (2016)	By applying Classification method, Predicting the students' performance based on related personal and social factors - It was slightly found that the student's performance is not totally dependent on their academic efforts
H. Hamsa, S. Indira Devi and J. Kizhakkethottam (2016)	By applying Classification method, Develop student's Academic performance prediction model- A decision to take care about risk students and mental satisfaction for safe students.
C. Anuradha and T. Velmurugan (2016)	By applying Feature Selection Classification method, Show that the feature selection techniques can improve the accuracy and efficiency of the Classification algorithm - Improve the student performance.
B. Rawat and S. Dwivedi (2017)	By applying Clustering, Recommendation of courses to a learner based on his/her profile - Predicting student's performance timely can help them to improve their learning process, Consequently improving student academic performance.
Sunil and M. Doja (2017)	By applying Association rules and Clustering with collaborative Filtering, Find the students visiting patterns - Construct a browsing behavioural model that is helpful in supporting E-learning resources.

IX. APPLICATIONS OF DATA MINING

Educational data mining is a multidisciplinary research area; hence it is difficult to bind it in few applications. But to list few of them, primary applications of EDM are [13]

- A. Predicting student performance
- B. Student modeling
- C. detecting undesirable student behaviors
- D. Analysis and visualization of data
- E. providing feedback for supporting instructors
- F. constructing courseware
- G. Planning and scheduling
- H. Recommendations for students
- I. grouping students
- J. Social network analysis
- K. Developing concept maps

X. CONCLUSIONS

This paper described the detailed study of Educational data mining, the objectives of the Educational data mining and the phases of the Educational data mining and the existing classification techniques. Educational data mining is a young research area. It is an emerging field related to several well-established research areas that include e-learning, adaptive hypermedia, smart tutoring systems, web mining, data mining, etc., mainly the need to take into account the pedagogical aspects of the learner and the system. Although the mining of educational data is a very recent area of research, there are a significant number of published contributions. EDM brings together researchers and professionals in informatics, education, psychometrics, statistics, psychology, etc.

REFERENCES

- [1] R.Jindal, M.D Borah, A Survey on Educational Data Mining and Research trends, International Journal of Database Management System (IJDMS), 5(3), 2013, 53–73.
- [2] N. Upadhyay, V. Katiyar, A Survey on the Classification Techniques in Educational Data Mining, International Journal of Database Management System (IJDMS), 3(11), 2014, 725–728.
- [3] Baradwaj, B. Kumar, Mining Educational Data to Analyze Students Performance. International Journal of Advanced Computer Science and Applications (IJACSA), 2(6), 2011, 63-69
- [4] Geeta Kashyap, Ekta Chauhan, REVIEW ON EDUCATIONAL DATA MINING TECHNIQUES, International Journal of Advanced Technology in Engineering and Science, vol.no.3, issue 11, November, 2015.
- [5] C. Romero and S. Ventura, "Data mining in education," WIREs Data Mining and Knowledge Discovery, vol. 3, pp.12-27, 2012
- [6] C. Papamitsiou and A. A. Economides, "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence," Educational Technology & Society, vol. 17, no. 4, pp. 49-64, 2014.
- [7] H. Kumar and N. Modi, "A Survey on Educational Data Mining and Research Trends," KAAV INTERNATIONAL JOURNAL OF SCIENCE, vol. 4, no. 3, pp. 84-89, 2017
- [8] Jaya Srivastava, Dr Abhay Kumar Srivastava, Data Mining in Education Sector: A Review, National Conference on Cloud Computing & Big Data.
- [9] Vandna Dahiya, A SURVEY ON EDUCATIONAL DATA MINING, International Journal of Research in Humanities, Arts and Literature (IMPACT: IJRHAL) ISSN (P): 2347-4564; ISSN (E): 2321-8878, Vol. 6, Issue 5, May 2018, 23-30.
- [10] Baker R S J d, Yacef K. The state of educational data mining in 2009: A review and future visions. J EduData Min 2009.
- [11] R. Jindal and M. Borah, "A Survey on Educational Data Mining and Research Trends", International Journal of Database Management Systems, vol. 5, no. 3, pp. 53-73, 2013.
- [12] C. Romero, S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Systems with Applications 33 (2007) 135–146 (Science Direct).
- [13] Mr.Suhas G. Kulkarni 1, Mr.Ganesh C. Rampure 2, Mr.Bhagwat Yadav, Understanding Educational Data Mining (EDM), International Journal of Electronics and Computer Science Engineering 773, ISSN 2277-1956/V2N2-773-777
- [14] Sunil and M. Doja2, "Data Mining Techniques to Discover Students Visiting Patterns in E-learning Resources", International Journal of Computer Science and Mobile Computing, vol. 6, no. 6, pp. 363 – 368, 2017.
- [15] S. Borkar and K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining", International Journal of Computer Science and Mobile Computing, vol. 2, no. 7, pp. 273 – 279, 2013.
- [16] B. Rawat and S. Dwivedi, "An Architecture for Recommendation of Courses in E-learning System", IJ. Information Technology and Computer Science, vol. 4, pp. 39-47, 2017.
- [17] M. Chakurkar and P. Adiga, "A Web Mining Approach for Personalized E-Learning System", International Journal of Advanced Computer Science and Applications, vol. 5, no. 3, pp. 51-56, 2014.
- [18] A. Abu, "Educational Data Mining & Students' Performance Prediction", International Journal of Advanced Computer Science and Applications, vol. 7, no. 5, pp. 212-220, 2016.



- [19] "Educational Data Mining and its role in Educational Field", International Journal of Computer Science and Information Technologies, vol. 5, no. 2, pp. 2458-2461, 2014.
- [20] I. P. Ratnapala, R. G. Ragel and S. Deegalla, "Students behavioural analysis in an online learning environment using data Mining," 7th International Conference on Information and Automation for Sustainability, Colombo, 2014, pp.1-7, doi: 10.1109/ICIAFS.2014.7069609.
- [21] T. Z. and A. M.Mahmoud, "Clustering of Slow Learners Behaviour for Discovery of Optimal Patterns of Learning", International Journal of Advanced Computer Science and Applications, vol. 5, no. 11, pp. 102-108, 2014.
- [22] A. Bogarín, C. Romero and R. Cerezo, "Discovering students' navigation paths in Moodle", in Proceedings of the 8th International Conference on Educational Data Mining, 2015.
- [23] C. Anuradha and T. Velmurugan, "FEATURE SELECTION TECHNIQUES TO ANALYSE STUDENT ACADAMIC PERFORMANCE USING NAÏVE BAYES CLASSIFIER", in The 3rd International Conference on Small & Medium Business 2016, Nikko Saigon Hotel, Hochiminh, Vietnam, 2016.
- [24] H. Hamsa, S. Indira Devi and J. Kizhakkethottam, "Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm", Procedia Technology, vol. 25, pp. 326-332, 2016.
- [25] K. Rangra and K. Bansal, "Comparative Study of Data Mining Tools", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 6, pp. 216-223, 2014.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)