iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35431

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: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Survey Paper on Information based Learning System using Educational Data Mining

Dhanendra Kumar[1], Dr. Rajesh Kumar Pathak[2], Mr. Komal Yadav[3]
[1] *Research Scholar CSE, Shri Rawatpura Sarkar University Raipur,Chhattisgarh India*
[2]*Professor, Vice Chancellor, Shri Rawatpura Sarkar University Raipur,Chhattisgarh India*
[3]*Assistant Professor, CSE, Shri Rawatpura Sarkar University Raipur,Chhattisgarh India*

Abstract: ***Educational Data Mining (EDM) is a platform for learning and exploring from data to get essential information and generate the unique pattern which will help study, analyse and skill student performance in academic. Various data mining methods can be apply to filter the data from data warehouse to implement data mining techniques which helps student for taking decisions for better outcome. The model which can be use in Educational data mining must be a constructive and descriptive model applied on data warehouse and must gather very accurate data for enhance the performance of study.***
Regression analysis can also be used to develop a model to use as study tool; it can be used dependent or independent variables. If the model is enough perfect for using as study tool then every cluster of data must be use that model to fetch the resultant data. Sometimes educational data mining is considered as overall performance of students, but each student has its own level of understanding the contents so that method must also be enough flexible for every one ; for fulfilling this requirement educational method can be complex, but once it is constructed then it will helpful for every one. This paper is describing various data mining techniques and their proper uses.
Keywords: Educational Data Mining, Cluster Analysis, Classification, Regression Model, K-Means.

I. INTRODUCTION

Data mining techniques are wont to extract useful knowledge from data . The extracted knowledge is effective and significantly affects the choice maker. Educational data processing (EDM) may be a method for extracting useful information that would potentially affect a corporation . the rise of technology use in educational systems has led to the storage of huge amounts of student data, which makes it important to use EDM to enhance teaching and learning processes. EDM is beneficial in many various areas including identifying at-risk students, identifying priority learning needs for various groups of scholars , increasing graduation rates, effectively assessing institutional performance, maximizing campus resources, and optimizing subject curriculum renewal. This paper surveys the relevant studies within the EDM field and includes the info and methodologies utilized in those studies. Over the past decade there has been a rapid climb in education system. tons of latest institutions have come up both from public and personal sector offering sort of courses for under graduating and post graduating students.

The rates of enrolments for education has also increased but not the maximum amount because the number of upper institutions are increasing. it's a priority for today's education system and this gap has got to be identified and properly addressed to the training community. Hence it's become important to know the need of scholars and their academic progression. Educational data processing helps during a big thanks to answer the problems of predictions and profiling of not only students but other stake holders of education sectors.

The main goal of this mapping study is to define and answer the research questions supported analyzed articles. After the conducted search the relevant papers are selected and classification scheme is defined. we've answered the research questions as results of the systematic map and outcome of the method .

It often provides a visible summary, a map, of its results. First, are collected all publications needed for the interested field. At an equivalent time an summary of this research area is provided, the number , sort of research and available results are identified. The second step is that the conduct look for primary studies, and excluded are the studies that aren't relevant to answer. Here the research inquiries to drive the structure are provided. The third step is ensuring that the scheme takes the prevailing studies under consideration and providing better results. Answering the research questions provides data extraction and complete mapping of studies, by identifying, analyzing and interpreting the relevant evidence. The classification scheme drives the sector of interests, video field of interest and technology utilized in FC.

©IJRASET: All Rights are Reserved

A. Cluster of Students

In this case groups of scholars are created consistent with their customized features, personal characteristics, etc. These clusters/groups of scholars are often employed by the instructor/developer to create a customized learning system which may promote effective group learning.

The DM techniques utilized in this task are classification and clustering. Different clustering algorithms that are wont to group students are hierarchical agglomerative clustering, K-means and model-based clustering. A clustering algorithm is predicated on large generalized sequences which help to seek out groups of scholars with similar learning characteristics like hierarchical clustering algorithm which are utilized in intelligent e-learning systems to group students consistent with their individual learning style preferences.

B. Registration Management

This term is usually utilized in education to explain well-planned strategies and tactics to shape the enrolment of an establishment and meet established goals.

Enrolment management is an organizational concept and a scientific set of activities designed to enable educational institutions to exert more influence over their student enrolments.

Such practices often include marketing, admission policies, retention programs, and aid awarding. Strategies and tactics are informed by collection, analysis, and use of knowledge to project successful outcomes. Activities that produce measurable improvements in yields are continued and/or expanded, while those activities that don't are discontinued or restructured. Competitive efforts to recruit students are a standard emphasis of enrolment managers.

C. Survey and Determination of Data

It is wont to highlight useful information and support deciding . within the educational environment, for instance , it can help educators and course administrators to research the students' course activities and usage information to urge a general view of a student's learning. Statistics and visualization information are the 2 main techniques that are most generally used for this task. Statistics may be a mathematical science concerning the gathering, analysis, interpretation or explanation, and presentation of knowledge. it's relatively easy to urge basic descriptive statistics from statistical software, like SPSS.

Statistical analysis of educational data (logs files/databases) can tell us things like where students enter and exit, the foremost popular pages students browse, number of downloads of e-learning resources, number of various pages browsed and total time for browsing different pages.

It also provides knowledge about usage summaries and reports on weekly and monthly user trends, amount of fabric students might undergo and therefore the order during which students study topics, patterns of studying activity, timing and sequencing of events, and therefore the content analysis of scholars notes and summaries.

Statistical analysis is additionally very useful to get reports assessing what percentage minutes student worked, number of problems here solved and his correct percentage alongside our prediction about his score and performance level. Visualization uses graphic techniques to assist people to know and analyse data.

D. Predicting Student Performance

In this case, we estimate the unknown value of a variable that describes the student. In education, the values normally predicted are student's performance, their knowledge, score, or marks. This value can be numerical/continuous (regression task) or categorical/discrete (classification task). Regression analysis is used to find relation between a dependent variable and one or more independent variables. Classification is used to group individual items based upon quantitative characteristics inherent in the items or on training set of previously labelled items.

Prediction of a student's performance is the most popular applications of DM in education. Different techniques and models are applied like neural networks, Bayesian networks, rule based systems, regression, and correlation analysis to analyze educational data. This analysis helps us to predict student's performance i.e. to predict about his success in a course and to predict about his final grade based on features extracted from logged data. Different types of rule-based systems have been applied to predict student's performance (mark prediction) in an learning environment (using fuzzy-association rules).Several regression techniques are used to predict student's marks like linear regression for predicting student's academic performance, stepwise linear regression for predicting time to be spent on a learning page, multiple linear regression for identifying variables that could predict success in colleges courses and for predicting exam results in distance education courses.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

II. BACKGROUND AND RELATED WORK

In this section we found that many authors have tried to find out the technique by which the educational data can take filtration at optimum level. Various clusters can be created by the researchers so that the educational data must be simplifies in proper manner. They also used various algorithms to predict the accurate data.

It helps to identify the students' performance range like average, below average, and good performance. As there are several approaches that area unit used for knowledge classification. . This study can facilitate the students and the lecturers to boost the students of all category to perform well.

J K Jothi and K Venkatalakshmi conducted the students' performance analysis on the graduate students' data collected from the Villupuram college of Engineering and Technology.

The data included five year period and applied clustering methods on the data to overcome the problem of low score of graduate students, and to raise students academic performance.[1]

Sheik and Gadage have done the analysis related to the student learning behavior by using different data mining models, namely classification, clustering, decision tree, sequential pattern mining and text mining. They used open source tools such as KNIME (Konstanz Information Miner), RAPIDMINER, WEKA, CARROT, ORANGE, RProgramming, and iDA. These tools have different compatibilities and it provided an insight into the prediction and evaluvation.[2]

Mythili M S and Shanavas A R applied classification algorithms to analyze and evaluate school students' performance using weka. They came with various classification algorithms, namely J48, Random Forest,

Multilayer perception, IBI and decision table with the data collected from the student management system [3].

Dinesh A and Radhika V targeted on the techniques and strategies of instructional data processing for data discovery from the information collected from various universities. This paper stated that relationship mining was leading between 1995 and 2005 and in 2008 to 2009 it slipped to 5th place. During the period 2008 to 2015 45% papers are moving to prediction. The prediction model acts like a warning system to improve their performance [4].

Osmanbegovic and Suljic conducted a study for investigating students' future performance in the end semester results at the University of Tuzla. They considered 11 factors and used classification model with highest accuracy for naive Bayes [5].

Suyal and Mohod applied the association and classification rule to identify the students' performance. They mainly focused to find the students who need special attention to reduce failure rate [6].

Noah, Barida and Egerton conducted a study to evaluate students' performance by grouping the grading into various classes using CGPA. They used different methods like Neural network, Regression and K-means to identify the weak performers for the purpose of performance improvement. The prediction with high accuracy in students' performance is beneficial as it helps in identifying the students with low academic achievements at the early stage of acdemics. In universities, student retention is related to academic performance and enrollment system. [7].

Baradwaj and pal described data mining techniques that help in early identification of student dropouts and students who need special attention. Here they used a decision tree by using information like attendance, class test, semester and assignment marks [8].

Jeevalatha, Ananthi, and Saravana Kumar presented a case study on performance analysis for placement selection for undergraduate students. They applied decision tree algorithm by considering the factors like HSC, UG marks and communication skills [9].

Backer and Yacef conducted a study for identifying the most appropriate model for EDM. They analyzed data and reached the conclusion that most of the papers adopt prediction than relationship mining [10].

ElGamal A F presented a study for predicting student performance in a programming course. Here the data is collected from the department of computer science from Mansoura University and applied extract rules for predicting students' performance in programming course [11].

Angeline D M conducted a study on the students' performance by using Apriori algorithm that extracts the set of rules specific to every category and analyze the given knowledge to classify the scholar based on their involvement in assignment, internal assessment test, group action etc. It helps to identify the students' performance range like average, below average, and good performance [12].

Bhise, Thorat and Supekar presented a method using K-means clustering algorithm by describing it step by step. This paper mainly focused on reducing drop-out-ratio of the students and improve it by considering the valuation factors like midterm and final exam assignment.

They considered different clustering techniques namely hierarchical, partitions, and categorical. This study can facilitate the students and the lecturers to boost the students of all category to perform well. This study helps to spot out those students who require special attention , minimize the failure ratio and to take acceptable action for upcoming semester examination. [13].

©IJRASET: All Rights are Reserved

Remesh, Parkavi, and Yasodha conducted a study on the placement chance prediction by investigating the different techniques such as Naive Bayes Simple, MultiLayerPerception, SMO, J48, and REPTree by its accuracy. From the result they concluded that MultiLayerPerception technique is more suitable than other algorithms [14].

Tair M M A and El-Halees presented a case study with a set of data collected from degree holders of college 'Science and Technology, Khanyounis', during the period of 1993 to 2007.

They used two classification methodologies such as Rule Induction and Naive Bayesian classifier to forecast the grades of the students. the classification is employed in student information to predict the students' division on the premise of previous information.

As there are several approaches that area unit used for knowledge classification, Naive theorem is employed here. Information like group action, class test, seminar and assignment marks were collected from the students' previous information, to predict the performance at the top of the semester. [15].

III.SURVEY METHODOLOGY

This research performed a scientific review where the relevant academic works predicting student performance using learning output is identified, selected, and critically evaluated using several criteria, as presented within the results section. To stream line our contributions, we formulated three key research questions as follows:

Learning Outcomes Prediction. How is student academic performance measured using learning outcomes?

Academic Performance Prediction Approaches. What intelligent models and techniques are devised to forecast student academic performance using learning outcomes?

Academic Performance Predictors. What dominant predictors of student performance using learning outcomes are reported?

The main objective of this survey was to create a comprehensive understanding of the landscape of academic performance prediction by focusing on the attainment of learning outcomes. To achieve the performance of students we simulate row data and after processing we find out the results.

A. Algorithm Prototype

The proposed algorithm will complete in following process:

1) First of all raw data must be collected. The data will be huge due to inter related various homogeneous data.So the data collection is important for further process.

2) Second step is Implementation of the classification scheme
Classification scheme can be implemented to divide the data in various data clusters.

3) Associate the info query in standard format Queries are going to be formed within the mode of standardisation. Which involves the hearth within the physical data barriers.

4) Implement the prediction function for Data Comparison The method will compare the particular data with predictive analytical data for choosing accurate data.

5) Time estimation for analysis Time analysis is going to be a serious part where retrieval methods can compare the minimum time for extracting data.

6) Create a cluster of knowledge for predictive result Cluster of knowledge will help to form sure that data comes from physical section is most accurate.

7) Impose the clean method for extract the info The method which has responsibility to pick most accurate cluster for many accurate result.

8) Examine the Result by comparing output Resultant data are often found using comparative method with most effective cluster and fewer time counter. This algorithm will find out the most relevant data for final output because irrelevant data will be escaped by comparing data in data ware house. Final output will produce after applying all processes available in algorithm.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

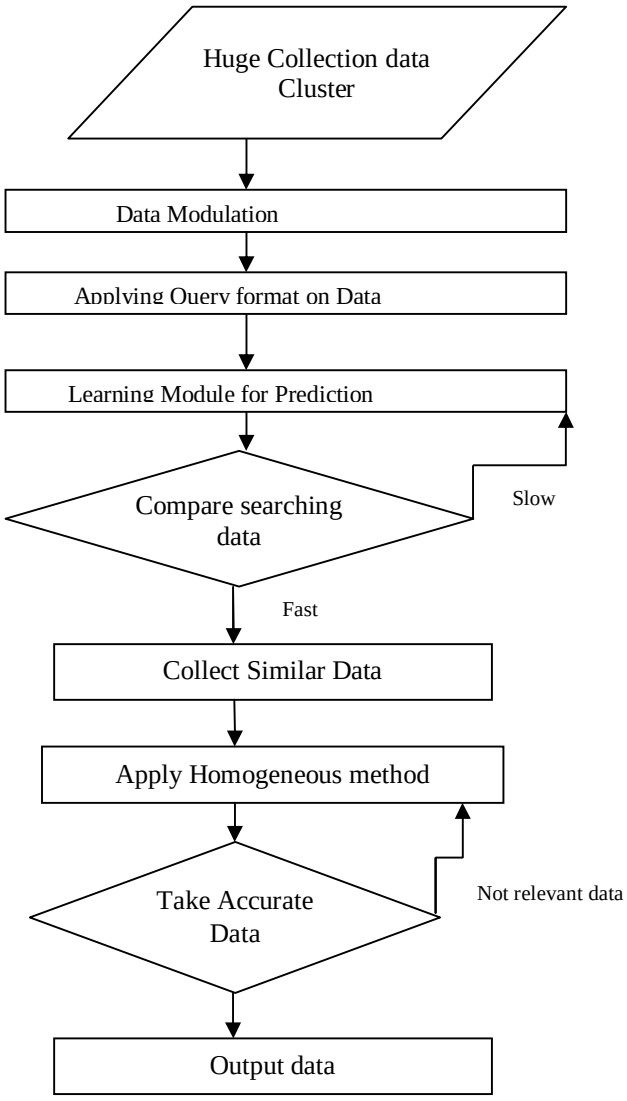


Fig 1: Predictive Based flow chart

IV. RESULT AND ANALYSIS

In this section, explain the varied ways during which performance of scholars are often studied, one among the ways being Correlation. The algorithm will help us to seek out the will data in accurate manner. Decision tree algorithm under Classification technique are often wont to find the result. It predicts the searching data for educational data processing . Educational data processing (EDM) describes a search field concerned with the appliance of knowledge mining, machine learning and statistics to information generated from educational settings.

V. CONCLUSION

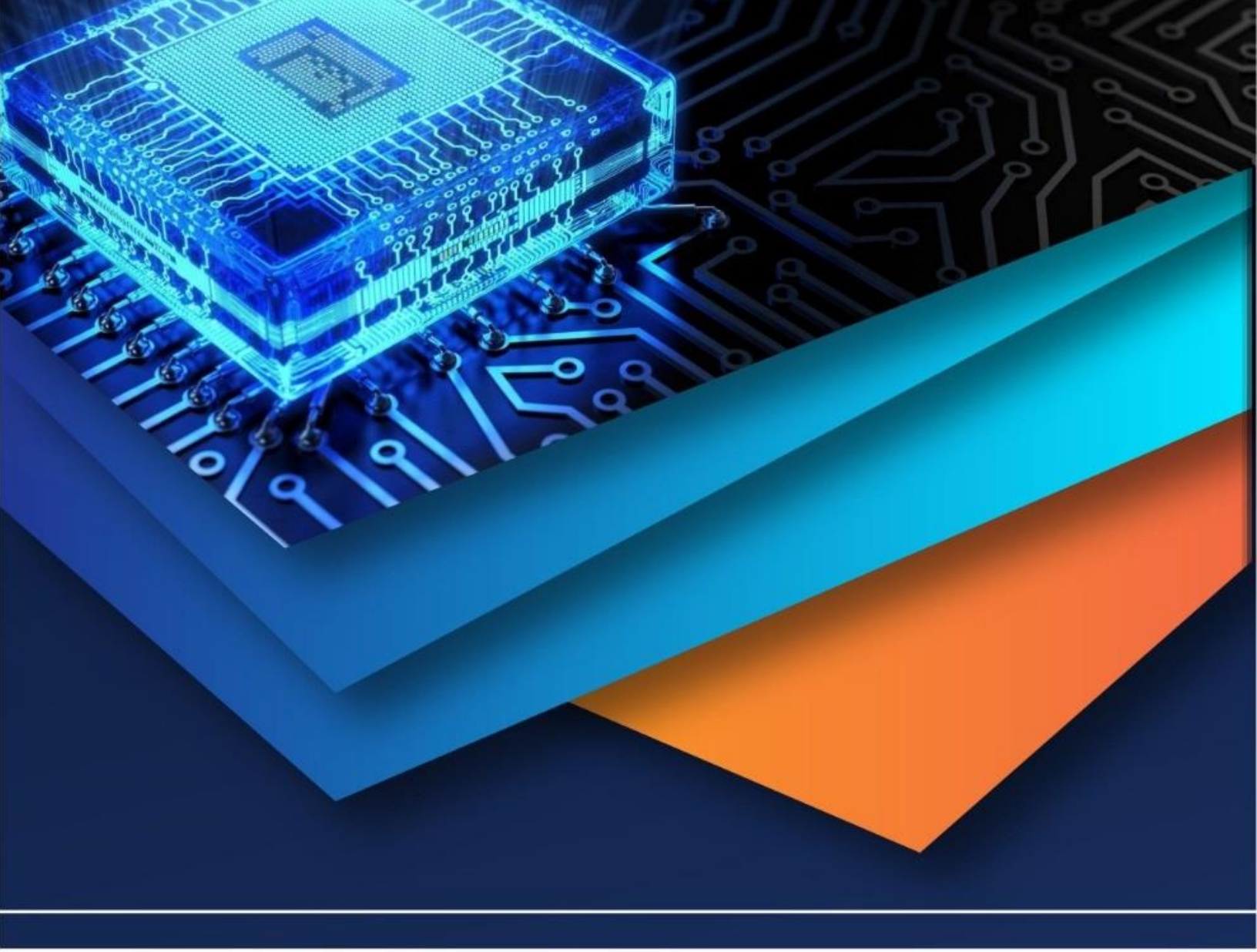
In this paper, Educational data processing is mentioned the method of extracting hidden and useful information in large data repositories. Knowledge Discovery and data processing (KDD) may be a multidisciplinary area focusing upon methodologies for extracting useful knowledge from data and there are several useful KDD tools to extracting the knowledge. this data are often wont to increase the standard of education. Educational data processing cares with developing new methods to get knowledge from educational/academic database and may be used for deciding in educational/academic systems. This paper discusses about what's educational data processing , its broad application areas, benefits of educational data processing , challenges and barriers to successful application of educational data processing and therefore the new practices that need to be adopted so as to successfully employ educational data processing and learning analytics for improving teaching and learning.

©IJRASET: All Rights are Reserved

REFERENCES

[1] J.K.Jothi and K.Venkatalakshmi, “Intellectual performance analysis of students by using data mining techniques”, International Journal of Innovative Research in Science, Engineering and Technology, vol 3, Special iss 3, March 2014.
[2] Nikitaben Shelke and Shriniwas Gadage, “A survey of data mining approaches in performance analysis and evaluation”, International Journal of Advanced Research in Computer Science and Software Engineering , vol 5, iss 4, 2015K.
[3] M.S. Mythili1 and A.R.Mohamed Shanavas , “An analysis of students’ Performance using classification algorithms ”, IOSR-JCE, Volume 16, iss1, Jan. 2014.
[4] A.Dinesh Kumar and V.Radhika, “A survey on predicting student performance”, International Journal of Computer Science and Information Technologies, Vol. 5, 2014.
[5] E. Osmanbegović and M. Suljić, ”’ata mining approach for predicting students performance”, Economic Review, vol 10, iss 1, 2012.
[6] Sayali Rajesh Suyal and Mohini Mukund Mohod, “Quality improvisation of student performance using data mining techniques”, International Journal of Scientific and Research Publications, vol 4,iss 4, April 2014.
[7] OTOBO Firstman Noah, BAAH Barida and Taylor Onate Egerton, “Evaluation of student performance using data mining over a given data space”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, iss 4, September 2013.
[8] Brijesh Kumar Baradwaj and Saurabh Pal, “Mining educational data to analyze Ssudents’ performance”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.
[9] T.Jeevalatha, N. Ananthi and D.Saravana Kumar, “Performance analysis of undergraduate students placement selection using Decision Tree Algorithms”, International Journal of Computer Applications (0975- 8887), vol 108, December 2012.
[10] Ryan J.D.B.Baker and Kalina Yacef , “The state of educational data mining in 2009: A review and future revisions”, Journal of Educational Data Mining , Vol.1,No.1, February 2009.

©IJRASET: All Rights are Reserved



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)