iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** II **Month of publication:** February 2022

DOI: https://doi.org/10.22214/ijraset.2022.40306

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: 7.538
Volume 10 Issue II Feb 2022- Available at www.ijraset.com

Survey Paper on Enhancing Information Based Learning System Using Educational Data Mining

Mukesh Chandra[1], Shital Gupta[2]
[1]*Research Scholar,* [2]*Associate Professor, CSE, School of Research and Technology, People's University, Bhopal (M.P.) India*

Abstract: ***Educational Data Mining (EDM) is a platform for literacy and exploring from data to get essential information and induce the unique pattern which will help study, assay and skill pupil performance in academic. Colourful data mining styles can be apply to sludge the data from data storehouse to apply data mining ways which helps pupil for taking opinions for better outgrowth. The model which can be use in Educational data mining must be a formative and descriptive model applied on data storehouse and must gather veritably accurate data for enhance the performance of study.***
Retrogression 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 also every cluster of data must be use that model to cost the attendant data. Occasionally educational data mining is considered as overall performance of scholars, but each pupil has its own position of understanding the contents so that system must also be enough flexible for every one; for fulfilling this demand educational system can be complex, but once it's constructed also it'll helpful for everyone. This paper is describing colourful data mining ways and their proper uses.
Keywords: Educational Data Mining, Cluster Analysis, Classification, Regression Model, K-Means.

I. INTRODUCTION

Data booby-trapping ways are used to prize useful knowledge from data. The uprooted knowledge is effective and significantly affects the choice maker. Educational data processing (EDM) may be a system for rooting useful information that would potentially affect a pot. the rise of technology use in educational systems has led to the storehouse of huge quantities of pupil data, which makes it important to use EDM to enhance tutoring and literacy processes. EDM is salutary in numerous colorful areas including relating at- threat scholars, relating precedence literacy needs for colorful groups of scholars, adding scale rates, effectively assessing institutional performance, maximizing lot coffers, and optimizing subject class renewal. This paper surveys the applicable studies within the EDM field and includes the word and methodologies employed in those studies. Over the once decade there has been a rapid-fire rise in education system. Tons of rearmost institutions have come up both from public and particular sector offering kind of courses for under graduating and post graduating scholars.

A. Cluster of Students

In this case groups of scholars are created harmonious with their customized features, particular characteristics, etc. These clusters/ groups of scholars are frequently employed by the educator/ inventor to produce a customized literacy system which may promote effective group literacy. The DM ways employed in this task are bracket and clustering. Different clustering algorithms that are used to group scholars are hierarchical agglomerative clustering, K- means and model- grounded clustering. A clustering algorithm is rested on large generalized sequences which help to seek out groups of scholars with analogous literacy characteristics like hierarchical clustering algorithm which are employed in intelligent-learning systems to group scholars harmonious with their individual literacy style preferences.

B. Registration Management

This term is generally employed in education to explain well- planned strategies and tactics to shape the registration of an establishment and meet established pretensions. Registration operation is an organizational conception and a scientific set of conditioning designed to enable educational institutions to ply further influence over their pupil enrollments. Similar practices frequently include marketing, admission programs, retention programs, and aid awarding. Strategies and tactics are informed by collection, analysis, and use of knowledge to project successful issues. Conditioning that produce measurable advancements in yields are continued and/ or expanded, while those conditioning that do not are discontinued or restructured. Competitive sweats to retain scholars are a standard emphasis of registration directors.

©IJRASET: All Rights are Reserved

C. Survey and Determination of Data

It's used to punctuate useful information and support deciding. within the educational terrain, for case, it can help preceptors and course directors to probe the scholars course conditioning and operation information to prompt a general view of a pupil's literacy. Statistics and visualization information are the 2 main ways that are most generally used for this task. Statistics may be a fine wisdom concerning the gathering, analysis, interpretation or explanation, and donation of knowledge. it's fairly easy to prompt introductory descriptive statistics from statistical software, like SPSS. Statistical analysis of educational data (logs lines/ databases) can tell us effects like where scholars enter and exit, the foremost popular runners scholars browse, number of downloads on-learning coffers, number of colorful runners browsed and total time for browsing different runners.

D. Predicting Student Performance

In this case, we estimate the unknown value of a variable that describes the pupil. In education, the values typically prognosticated are pupil's performance, their knowledge, score, or marks. This value can be numerical/ nonstop (retrogression task) or categorical/ separate (bracket task). Retrogression analysis is used to find relation between a dependent variable and one or further independent variables. Bracket is used to group individual particulars grounded upon quantitative characteristics essential in the particulars or on training set of preliminarily labelled particulars. Vaccination of a pupil's performance is the most popular operations of DM in education. Different ways and models are applied like neural networks, Bayesian networks, rule grounded systems, retrogression, and correlation analysis to dissect educational data. This analysis helps us to prognosticate pupil's performance i.e. to prognosticate about his success in a course and to prognosticate about his final grade grounded on features uprooted from logged data.

II. BACKGROUND AND RELATED WORK

In this section we plant that numerous authors have tried to find out the fashion by which the educational data can take filtration at optimum position. Colorful clusters can be created by the experimenters so that the educational data must be simplifies in proper manner. They also used colorful algorithms to prognosticate the accurate data. It helps to identify the scholars' performance range like average, below average, and good performance. As there are several approaches that area unit used for knowledge bracket.. This study can grease the scholars and the speakers to boost the scholars of all order 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, sem 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].

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. 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

In our methodology we will use algorithm to search the prophetic data. This algorithm will help us to find the desire data in accurate manner. Decision tree algorithm under Bracket fashion can be used to find the result. It predicts the searching data for educational data mining. The algorithm will as follows:

Algorithm Prototype:

The proposed algorithm will complete in following process:

A. Simulate The Raw Data For Analysis From Data Warehouse

In first step data collection will be done for further process. The data will co relates various homogeneous data with large amount of data.

B. Implementation Of The Classification Scheme (Cleaning)

Classification scheme can be implemented to divide the data in various data clusters.

C. Feature Selection To Reduce Data Complexity

Feature will be formed in the mode of standardisation. Which comes to the fire in the physical data barriers with lots of features?

D. Get The Classification Method (K-Mean) For Comparative Analysis

The k-mean classification method will create data clusters and compare the actual data with predictive analytical data for selecting accurate data.

E. Interpret The Resultant Data

Resultant data can be found using comparative method with most efficient cluster and less time counter.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

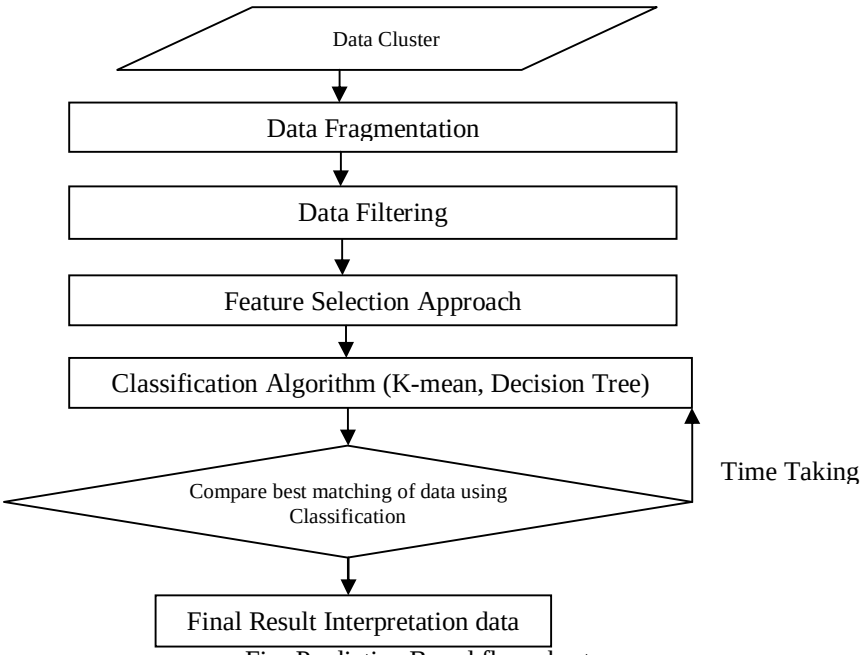


Fig: Predictive Based flow chart

IV. RESULT AND ANALYSIS

The affair of attendant data mining will be an accurate manner using the algorithm, algorithm filters the data using bracket approach. After bracket approach decision tree helps to get accurate result with Minimal redundancy. Mining Algorithm, information mining computations applied to set up and execute a model that finds and sums up information on enthusiasm to the customer (help, understudies and moderators). To do as similar, either broad or unequivocal information mining accoutrements or information mining instruments can be employed economically or for nothing.

Explain the colourful styles wherein performance of scholars may be studied, one of the ways being Correlation. The algorithm will help us to find the desire records in correct manner. Decision tree set of rules underneath Bracket approach can be used to find the result. It predicts the searching records for educational records mining. Educational Data Mining (EDM) describes a exploration field involved with the software of records mining, machine mastering and statistics to records generated from educational settings.

V. CONCLUSION

In this paper, Educational data processing is mentioned the system of rooting retired and useful information in large data depositories. Knowledge Discovery and data processing (KDD) may be a multidisciplinary area fastening upon methodologies for rooting useful knowledge from data and there are several useful KDD tools to rooting the knowledge. this data are frequently used to increase the standard of education. Educational data processing cares with developing new styles to get knowledge from educational/ academic database and may be used for deciding in educational/ academic systems. This paper discusses about what is educational data processing, its broad operation areas, benefits of educational data processing, challenges and walls to successful operation of educational data processing and thus the new practices that need to be espoused so as to successfully employ educational data processing and literacy analytics for perfecting tutoring and literacy.

REFERENCES

[1] Sara Fatima,Salma Mahgoub, " Predicting Student's Performance in Education using Data Mining Techniques", International Journal of Computer Applications (0975 – 8887), Volume 177 – No. 19, November 2019 .

[2] Sushil Shrestha, Manish Pokharel , " Educational data mining in moodle data", International Journal of Informatics and Communication Technology (IJ-ICT), Vol.10, No.1, April 2021, ISSN: 2252-8776.

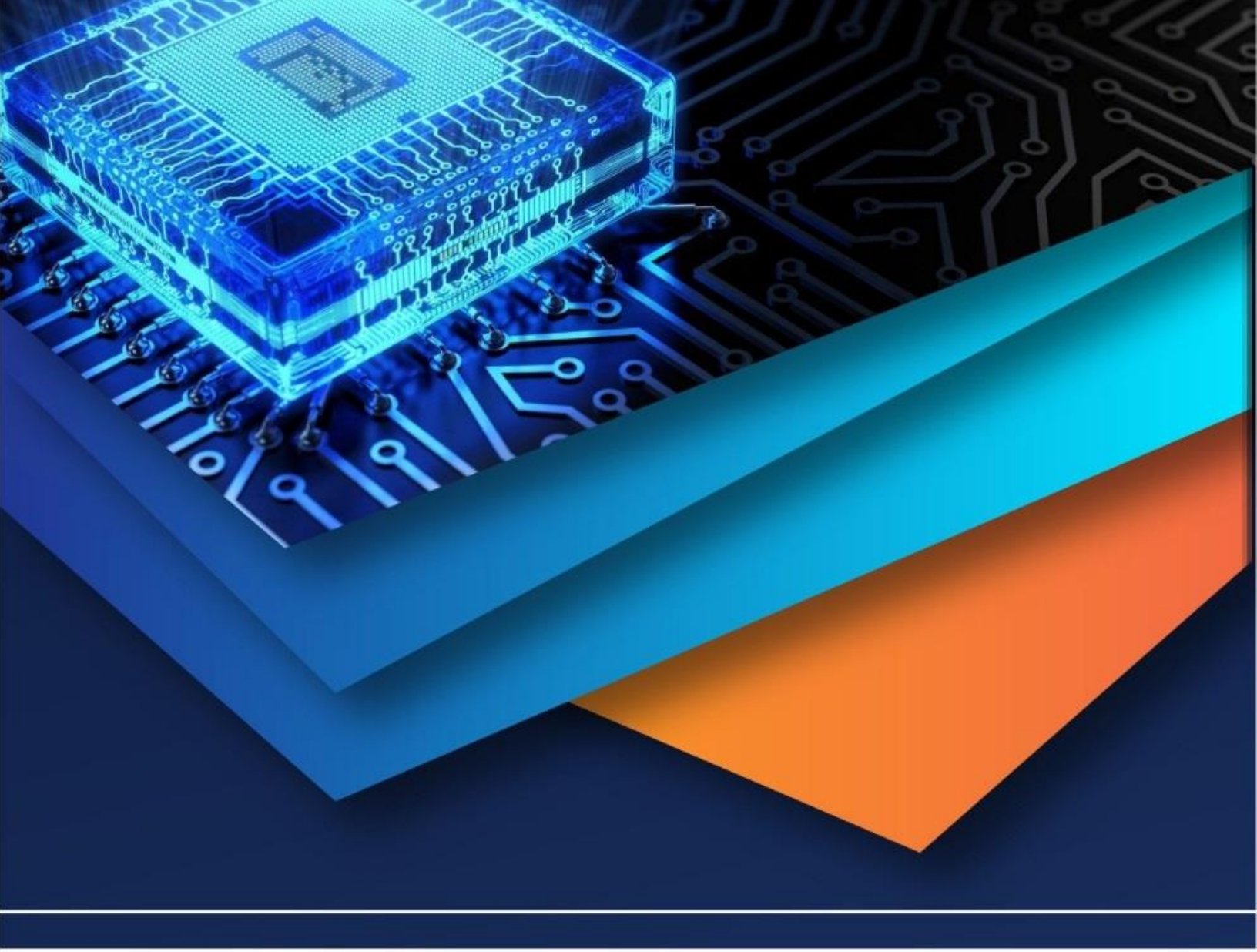
[3] Ahmed Saied Rahama Abdallah " Using Regression Analysis to Identify the Predictive Ability of the Achievement Test and the Secondary School Rate in the Prediction of the Cumulative Rate ", International Journal of Computer Applications (0975 – 8887), Volume 177 – No. 17, November 2019.

[4] Nouf S. Aldahwan, Nourah I. Alsaeed, " Use of Artificial Intelligent in Learning Management System (LMS): A Systematic Literature Review", International Journal of Computer Applications (0975 – 8887), Volume 175– No. 13, August 2020.

[5] Krishna Parmar, Huma Khan,"A Survey on Analysis the Students Mind in Different Area", International Journal of Science and Research (IJSR), ISSN: 2319-7064, Impact Factor (2017): 7.296.

[6] Nilesh V. Ingale, Dr. M. Sivakkumar, Dr. Varsha Namdeo , " Survey on Prediction System for Student Academic Performance using Educational Data Mining", Turkish Journal of Computer and Mathematics Education Vol.12 No.13 (2021), 363-369.
[7] Suleiman Khalifa Arafa Ibrahim, Mahmoud Ali Ahmed,"Prediction of Students' Cumulative Grade Point Averages (CGPAs) at Graduation: A Case Study" International Journal of Computer Applications (0975 – 8887), Volume 174 – No. 24, March 202.
[8] Nancy Kansal,Vineet Kansal, " An Efficient Data Mining Approach to Improve Students' Employability Prediction", International Journal of Computer Applications (0975 – 8887), Volume 178 – No. 47, September 2019.
[9] Sathyendranath Malli, Nagesh H. R.,B. Dinesh Rao, " Approximation to the K-Means Clustering Algorithm using PCA", International Journal of Computer Applications (0975 – 8887), Volume 175– No. 11, August 2020.
[10] Anirudhd Soni, Anansha Gupta,"Feature Selection for Performance Prediction using Decision Tree", International Journal of Computer Applications (0975 – 8887), Volume 183 – No. 17, July 2021.
[11] Fatima Alshareef, Hosam Alhakami, Tahani Alsubait, Abdullah Baz," Educational Data Mining Applications and Techniques", (IJACSA) International Journal of Advanced Computer Science and Applications,Vol. 11, No. 4, 2020.
[12] Hemlata Pate, Dr. Dhanraj Verma ," Performance Analysis of Feature Selection Techniques for Text Classification", International Research Journal on Advanced Science Hub (IRJASH), Volume 02 Issue 12S December 2020.
[13] Laura O. Moraes and Carlos Eduardo Pedreira ," Clustering Introductory Computer Science Exercises Using Topic Modelling Methods", Accepted Article. Published In Ieee Transactions On Learning Technologies (2021 IEEE.
[14] Chaman Verma, Zoltán Illés,Veronika Stoffová, Pradeep Kumar Singh ," Predicting Attitude of Indian student's towards ICT and Mobile Technology for Real-Time: Preliminary Results", DOI 10.1109/ACCESS.2020.3026934, IEEE.
[15] Miguel A. Prada, Manuel Domínguez," Educational data mining for tutoring support in higher education: A web-based tool case study in engineering degrees", DOI 10.1109/ACCESS.2020.3040858, IEEE.
[16] Smita Ghorpade, Seema Patil," Educational Data Mining: Tools And Techniques Study", 2020 IJRAR November 2020, Volume 7, Issue 4 E-ISSN 2348-1269, P- ISSN 2349-5138.
[17] Yijun Zhao, Qiangwen Xu," Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)".

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)