iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** X **Month of publication:** October 2020

DOI: https://doi.org/10.22214/ijraset.2020.31835

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 8 Issue X Oct 2020- Available at www.ijraset.com

Predictive Analytics of Education Data Based Learning Patterns - A Literature Review

Mamta Saxena
Research Scholar, MVN University
Assistant Professor, J B Knowledge Park Faridabad

Abstract: ***In recent years, data mining techniques have been implemented in different fields like marketing, science, tour & travel, electronic commerce but it also has an impact in the field of education. So mining of Educational data is known as Educational Data Mining. Nowadays, this field has emerged as a new trend in the market. It focuses on the learning patterns followed by students, their academic performance and their level of programming skills. To upgrade performance of students in academics and predicting their learning behavior and to retain their failure intensity Multi-Class Classification technique is applied.***

Keywords: Educational data mining; Data Mining Techniques Classification Algorithm

I. INTRODUCTION DATA MINING

In data mining data is available in voluminous form and a technique is required to find the useful form of data from it. For that we need some techniques and methods that help the database analysts to find the useful patterns that can be represented in the form of clusters, charts, graphs or some other pictorial form and can be used in analyzing the useful results. This all can be done by using the technique called as Data Mining. Data Mining is that technique which can be used to find the useful patterns and then to extract the data that can be beneficial to large number of users. Data Mining is the discovery of data that can be extracted, cleansed and selected from unusual patterns of data to find the most relevant attributes that can be selected as the relevant attribute for data mining.

The Knowledge discovery data process consists of following number of iterative steps to mine the data:

1) *Cleaning of Data:* To clean the duplicate and inconsistent data
2) *Integrate the Data:* Integrate the data collected from multiple sources
3) *Selection of Data:* To select the most relevant data for the applying various operations like aggregation, association
4) *Data Mining:* It is the process where various techniques are available to get the useful patterns for analysis of data
5) *Evaluation of Pattern:* To find the most relevant pattern
6) *Representation of Knowledge:* In this various techniques of knowledge representation are applied to give mined information to the users.

A. *Various types of Data mined in Data Mining*

Data mining techniques mine various types of data such as:

1) Data Warehouse
2) Transactional
3) Database
4) Sequential
5) Multimedia
6) Spatial
7) Order/Sequence

B. *Types of Patterns mined in Data Mining*

Data mining mines the following patterns such as:

1) Frequent Patterns, Associations and Correlations
2) Outlier Analysis
3) Discrimination and Characterization of Class/ Concept
4) Classification and Regression Analysis
5) Clustering

©IJRASET: All Rights are Reserved

C. Educational Data Mining

Data Mining is the process of finding out useful knowledge from the available data for which it used specific pattern extraction algorithm [15]. Data mining has its application in various fields like finance, biology, tourism, electronic commerce other than it also emerges as a boon in education sector and it results into a new field known as Educational Data Mining.[16]

Educational Data Mining refers to tools, techniques and research designed for extracting knowledge from large repository of data. EDM is a new trend in mining the data and finding useful patterns and knowledge for educational systems by using various educational attributes like enrollment number, course enrolled for , registration and by other useful parameters that comes from different levels like schools, colleges , universities etc. EDM is a new field to explore out. All the educational institutes also use data mining tools to explore and analyze student performance to predict their results to improve their performance and to keep them away from failure. The main goal of EDM is to improve process of learning and extract useful knowledge. It converts raw educational data into useful information for designing a system for learning and pattern extraction.

The various parts of EDM applications are: [3]

1) Analyze and visualize data
2) Provide feedback to the instructors
3) Provide suggestions to the students
4) Anticipate Student performance
5) Detecting undesirable behavior of students
6) Categorize Student

D. Data Mining Techniques [12][15]

Following techniques of data mining has been already used for various predictions in different fields like Education, Medical, and Research etc:

1) Prediction,
2) Classification ,
3) Neural Networks,
4) Pattern Recognition ,
5) Decision Trees ,
6) Naive – Bayes ,
7) K-Nearest Neighbor

Researcher's in this field focused on finding out useful knowledge to help educational Institutions to manage students in a better way as well as balance the level of education and in addition to it enhance the performance of their students.

Academic performance of Students also depends on various numbers of factors like personal, social- economic, psychological and other environmental variables. Prediction model uses all these factors to predict the student performance with high accuracy and those with low academic achievements. These predictions help the learners to upgrade performance of students.

E. Challenges that Occur in Mining Educational Data

1) *Dealing with Large Volume of Unstructured data and Different Types:* In EDM, data comes from various educational systems like Universities, Colleges, and Institutes in large amount and we need to find out the data useful for us.
2) *Extracting the Useful Patterns:* In EDM, it is a key challenge to extract the useful patterns form large voluminous data that are available from various educational systems and used that extracted data for the purpose it is required for
3) *Fitting of Data in Databases that are Comparatively in Small Size:* Since data that are available is large in volume but the size of database where we try to put our data for further use is small so it is a big challenge to fit the data according to the size of data.
4) *Prediction of Future Behavior of Students for the Learning Outcomes and to keep Failure Intensity Low:* To process the data in EDM for analyzing the future behavior of students from their learning outcomes which is beneficial for educators, learners, administrators and stakeholders? Though the task is difficult to predict about the future behavior of students as it is uncertain we cannot be assured about the future behavior of students.
5) *Description of some of the Attributes of Dataset:* Various attributes used for evaluating student's academic performance are [11]

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue X Oct 2020- Available at www.ijraset.com

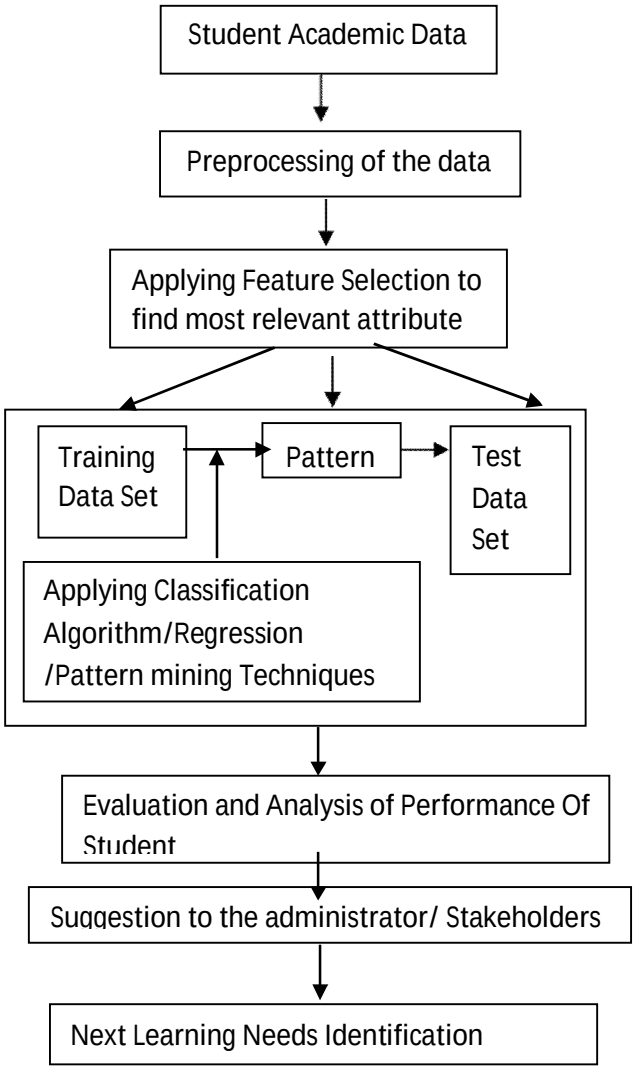
F. Attributes and their Description

G Gender of student whether Male or Female;
CAT Category of student whether General, SC, ST, OBC;
TNP Class Xth Percentage; TWP Class XIIth Percentage
IP Internal Assessment Percentage; EP End Semester Percentage; ATTEN Class attendance
BACKP Back in any Paper; MS Marital Status;
FAS Family Size;

G. Classification Algorithms

Classification algorithms describe about the performance of students and offer parameters to educators and learners that best describe about the performance of students and finding out the parameters needed by the educational settings. [11]

In the current educational system academic performance of students is determined by both assessments i.e. internal assessment as well as by the end semester examination system. Internal assessment consists of various attributes like Class test, seminars, attendance, projects , lab works etc. and on the basis of performance in the following parameters a teacher assess a student while in the end semester examination system students only need to score minimum marks to just pass the exams.



Framework for Student's Academic Performance

OF Occupation of Father;

OM Occupation of Mother

©IJRASET: All Rights are Reserved

H. Applying Feature Selection Method

This method describes about the most influential attribute used for classification.

After applying Feature Selection method, classification algorithms are applied [2] .There are various algorithms that can be used to check the academic performance of students. These algorithms are J48, PART, Bayes Net and Random Forest.

I. Classification Analysis

All these algorithms namely J48, BayesNet, PART and Random Forest are part of supervised learning algorithms that use training facts to test the correctness of testing data [12].

1) *J48 Classifier:* J48 classifier generates decision tree based on C4.5 algorithm. [12].
2) *BayesNet Classifier:* BayesNet classifier algorithm provides higher accuracy on large database. It also takes less computational time than Bayesian Network and gives result on the basis of conditional dependencies using direct graph [12].
3) *Random Forest Classifier:* Random Forest classifying algorithm uses a method known as bootstrap sampling method on the training dataset to build many unpruned classification trees. During the testing phase, mean of all unpruned classification trees provides the final predicted output for a randomly selected feature [13].
4) *PART Classifier:* It is a rule learning classifying algorithm that combines divide-and-conquer method with separate-and conquers strategy. It builds a partial decision tree on the current set of instances and creates a rule from the decision tree [14].

II. BRIEF SURVEY OF EARLIER RELATED WORK

The first aim of Educational data mining is to find out those students who need special attention to reduce their failure ratio and get promoted for next semester exams. It can be achieved by using various educational parameters as discussed in [1].

EDM focuses on generating classification rules and predicting student's academic performance in a course selected and previously recorded behavior and activities performed by the students. [2]

Research conducted on Educational Data Mining by using Naïve Bayes classification algorithm to analyze, classify, and predict students whether they are performers or underperformers? And it was helpful in classifying and predicting the grades of students for the upcoming year at an extent level.[3]

IN EDM Questionnaire method which had number of questions related to personal , social and psychological factors with parameters like student's performance and grades is also helpful in generating predictive variables and find out different factors that affect the behavior and performance of students while learning and validating their data by comparing students from different Universities and Institutions.[4]

Academic performance of students can also be analyzed by performing comparative research to test multiple decision tree algorithms on an educational dataset that classify the academic performance of students. [5]

The CART (Classification and Regression Tree) method works better on the tested dataset, which selects data based on the produced accuracy and precision using 10-fold cross validation. It showed a good practice of finding out the best classification algorithm technique for a selected dataset; that is by testing multiple algorithms and techniques before deciding which one will work better for the dataset and to test the dataset with multiple classifiers first, then choose the most accurate and precise one to decide the best classification method for any dataset.[5]

In order to analyze the learning performance of students another clustering algorithm named as K- Means Clustering Algorithm is used to predict learning activities of students which can be helpful for both educators and learners.

A research conducted about the study in private tutoring of students and it was observed that the total percentage of students those are taking private tuitions in India is comparatively higher than other countries namely Malaysia, Singapore, Japan, China and Sri Lanka. It was also stated that there was a number of repetitive increment in the academic performance of students with intensity of private tutoring and this sudden change of intensity of private tutoring depends on the factor likely socio economic conditions.[7]

Attributes like Mother's education and student's family financial gain also are related to with the opposite factors explicit like "Attitude of scholars towards attending at school, hours spent in study on routine day} after faculty, students' family financial gain, mother's age. This analysis was conducted on the sample of three hundred students.[8]

The case study uses student knowledge to investigate the learning behavior to predict the results and to alert students regarding any risk before their final exams.[9]

Decision tree model is additionally wont to predict the ultimate grades of scholars with alternative classification strategies like ID3, C4.5 and Naïve Thomas Bayes. [10]

©IJRASET: All Rights are Reserved

A. Outline of Literature Survey

To predict performance of the scholars based mostly upon their lecturers, knowledge is collected per completely different aspects of student's record as well as previous records of their educational performance, family background and alternative demographics. 3 classifier algorithms viz. call Tree, Naïve Thomas Bayes and Rule – based mostly classifier ar applied to search out the educational performance of scholars and their level of success.[11]

A study conducts to search out out extremely influenced attributes that predict regarding student performance. In such case Bayesian Network is useful find the accuracy rate per the student's attending and grade point average. Feature choice technique is used before any classification for such job however the Bayesian Network performs higher than the choice Tree classification during this case.

Classification models by victimisation call tree and artificial neural network techniques may additionally engineered to urge the strength and weakness of the scholars and to boost the student's performances.

WEKA tool is additionally useful in evaluating performance of the scholars with the accuracy of the classifier algorithms depends upon size and nature of information. From the studies it's clear that Bayesian Network performs higher than the opposite techniques in terms of accuracy.

A qualitative model was conjointly steered by Amjad Abu Saa to investigate the scholar performance supported students' personal and social factors [2]. The author devised numerous factors of the students' performance within the field of upper education that was a in theory conception.

III. GAPS WITHIN THE GIFT ANALYSIS

While rummaging the literature survey following gap in current analysis has been found Applying Classification algorithmic program in predicting the educational performance of scholars leads to confusion matrix. the answer to resolve downside of Confusion matrix isn't utterly resolved by victimisation numerous parameters like sensitivity, accuracy, exactitude and mean absolute error.

Decision tree models ar useful to classify students in numerous teams of fine, Average and dangerous and to not classify on the premise of their learning programming skills therefore clump Algorithms ar used.

Association Rule Mining could be a answer to search out out attention-grabbing patterns from knowledge of scholars rather than victimisation prophetical models to predict regarding the student's grades from collected coaching dataset. [9]

ID3 call algorithmic program is wont to build the choice tree by following top-down approach to scale back the frequency of queries asked whereas learning to avoid learner's distraction. If the algorithmic program apply greedy approach to optimize the performance of algorithmic program the performance is improved.[10]

IV. DOWNSIDE FORMULATION

Solution of the subsequent analysis downside is:

A. Classification techniques to investigate the training pattern of scholars.
B. Feature choice technique to search out the foremost relevant attributes.
C. Algorithms on analyzing Programming skills of scholars, learning patterns , analyzing educational performance
D. Implement the prevailing algorithms and planned algorithms and obtain the results.
E. Analyze the results of planned algorithms and check their performance with existing algorithms.

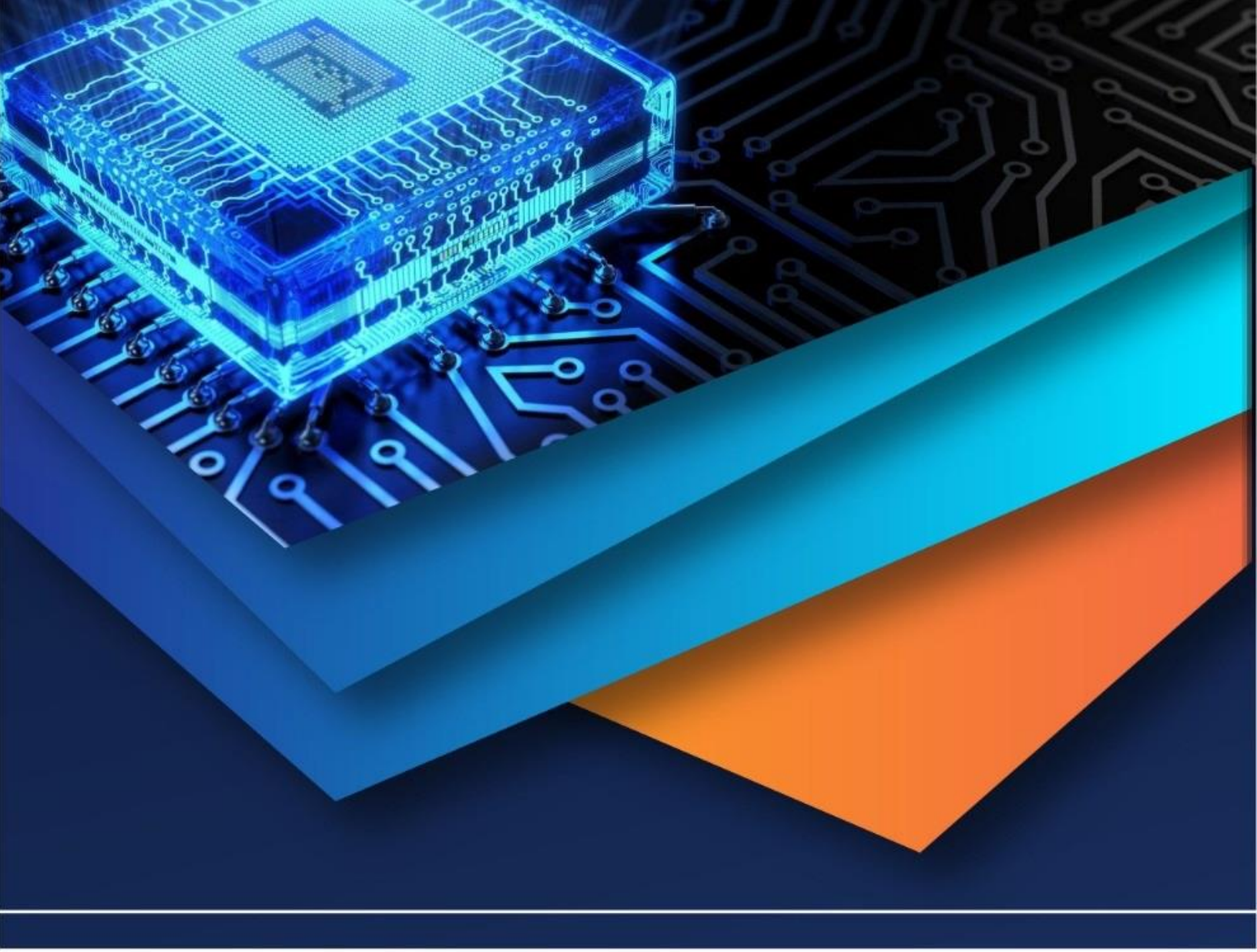
REFERENCES

[1] Bharadwaj, B.K. and Pal, S., 2011. Mining academic knowledge to investigate Students' Performance. (IJACSA) International Journal of Advanced engineering science and Applications, Vol. 2, No. 6, 2011.
[2] Ahmed, A.B.E.D. and Elaraby, I.S., 2014. knowledge Mining: A prediction for Student's Performance victimisation Classification technique. World Journal of pc Application and Technology, 2(2), pp.43-47.
[3] Pandey, U.K. and Pal, S., 2011. knowledge Mining: A prediction of performing artist or underperformer victimisation classification. (IJCSIT) International Journal of engineering science and knowledge Technologies, Vol. 2 (2), 2011, 686690.
[4] Bhardwaj, B.K. and Pal, S., 2012. knowledge Mining: A prediction for performance improvement victimisation classification. (IJCSIS) International Journal of engineering science and knowledge Security, Vol. 9, No. 4, April 2011.
[5] Yadav, S.K., Bharadwaj, B. and Pal, S., 2012. data processing Applications: A Comparative Study for Predicting Student's Performance. International Journal of Innovative Technology & inventive Engineering (ISSN: 2045-711), Vol. 1, No.12, December.
[6] Yadav, S.K. and Pal, S., 2012. knowledge mining: A prediction for performance improvement of engineering students victimisation classification. World of engineering science and knowledge Technology Journal (WCSIT). (ISSN: 2221-0741), Vol. 2, No. 2, 51- 56, 2012.
[7] Bray, M., & Kwok, P. (2003). Demand for personal supplementary tutoring: abstract issues, and socio- economic patterns in city. economic science of Education Review, 22(6), 611-620.(PDF) The effectiveness of personal tutoring: Students' perceptions as compared with thought schooling in Hong Kong.
[8] Naqvi, Sayyed. (2006). Factors moving students' performance A Case of personal faculties. Bangla Desh e-Journal of social science. 3.

©IJRASET: All Rights are Reserved

[9] Kumar Verma, Sushil & Thakur, R.S. (2017). Fuzzy Association Rule Mining based mostly Model to Predict Students' Performance. International Journal of Electrical and pc Engineering (IJECE). 7. 2223. 10.11591/ijece.v7i4.pp2223-2231.
[10] Al-Radaideh, Qasem & Al-Shawakfa, Emad & I. Al-Najjar, Mustafa. (2006). Mining Student knowledge victimisation call Trees. The International Arab Journal of knowledge Technology - IAJIT.
[11] Hussain, Sadiq & Abdulaziz Dahan, Neama & Mutaher Ba- Alwib, Fadl & Najoua, Ribata. (2018). academic data processing and Analysis of Students' educational Performance victimisation maori hen. Indonesian Journal of technology and engineering science. 9. 447-459. 10.11591/ijeecs.v9.i2.pp447-459.
[12] Almarabeh, H., Analysis of Students' Performance by victimisation completely different data processing Classifiers. I.J. trendy Education and engineering science, 2017. 9(8): p. 9-15.
[13] Agrawal, S., S.K. Vishwakarma, and A.K. Sharma, pattern process Classifier for Predicting Student's Performance in UG level. International Journal of notebook computer Applications, 2017. 172(8): p. 6.
[14] Tan, P.-N., M. Steinbach, and V. Kumar, CHAPER 5: Rule- primarily based Classifiers, in Introduction to process, A. Nordman, Editor. 2005, PEARSON EDUCATION: US.
[15] Fayyad, U. "From process to information Discovery in Databases." AiMagazine seventeen.3(1996):37-54.
[16] Anjewierden , Anjo , B.Kolloffel, and C.Hulshof. "Towards tutorial information processing: pattern info mining ways in which for machine-controlled chat analysis to understand and support inquiry learning processes." International Workshop on Applying process in e-Learning (ADML 2007) (2007).
[17] R. Eberhart, and J. Kennedy, (1995) a modern Optimizer pattern Particles Swarm Theory, Proc. Sixth International conference on little Machine and Human Science (Nagoya, Japan), IEEE Service Center, Piscataway, NJ, pp. 39-43.
[18] J. Kennedy, and R Eberhart, (1995), Particle Swarm improvement, IEEE Conference on Neural Networks, pp. 1942-1948, (Perth, Australia), Piscataway, NJ, IV, 1995.
[19] J. Kennedy and R. Eberhart. Swarm Intelligence. Morgan Kaufmann Publishers, Inc., metropolis, CA, 2001.

©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)