iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** IV **Month of publication:** April 2015

DOI:

www.ijraset.com

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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Enhancement of educational system using data mining techniques

Prof.Priya Thakarei[1], Ajinkya Kunjir[2], Poonam Pardeshi[3], Shrinik Dosh[4], Karan Naik[5]
[1,2,3,4,5]*Department of Computer Engineering*
[1,2,3,4,5]*Sinhgad Institute of Technology and Science, Pune, India*

Abstract-In this paper we will discuss about the problem that are faced by education institutions. One of the biggest challenges that education faces today is predicting the right path of students. Institutions would like to know, which students will enroll in which course, and which students will need more assistance in particular subject and what efforts should be taken for weak students. Also it will help to predict the electives of the students and predict the dropout rate. Also some time management needs more information about student like their overall result, interest in co-curricular and about the success ofnew offered courses. One way to effectively address the challenges for improving the quality of students and education is to provide new knowledge related to the educational processes and entities to the system. This knowledge can be extracted from historical data that reside in the educational organization's databases using the techniques of data mining technology. If data mining techniques such as clustering, decision tree, association, classification and prediction can be applied to higher education processes, it can definitely help improve students' overall performance, their life cycle management, selection of course and predict their dropout rate.
Keywords: Data mining, Higher education, Clustering, Decision tree, neural network, classification, prediction, association rule analysis.

I. INTRODUCTION

One of the significant facts in learning institution is the explosive increase of educational data. These data are increasing rapidly without any benefit to the management and institutions.[1]We believe that to manage this vast data is difficult task, but by new techniques and tools we can easily process the large amount of generated data in business processes and extract some useful knowledge and information from it. Data mining is a technique of extraction hidden predictive information from large databases; it is a powerful new technology with great potential to help higher learning Universities or institutions to focus on the most important information in their data warehouses.[5] Data mining tools predict future trends and behavior patterns, allowing institution to make proactive, knowledge-driven and appropriate decisions. Data mining tools can answer institution questions that traditionally were too time consuming in past to resolve [4]. Education institutions can use classification technique, for a comprehensive analysis of student characteristics, or use estimation and prediction technique to predict the likelihood of a variety of outcomes, choosing elective, choosing right career path, drop out and course success.[6]

A. .Algorithms used

1) Naives Bayes algorithm

B. Tools of Data Collection & Analysis

Various tools are needed for project are for analyzing data, designing, implementation and some developing software tool such as:

1) Mysql database
2) Excel
3) Ms access
4) Weka data mining tool

©IJRASET 2015: All Rights are Reserved

Volume 3 Issue IV, April 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

II. DATA MINING: A WAY TO IMPROVE TODAYS HIGHER LEARNING INSTITUTIONS

Data mining is a powerful, new and emerging technology with great potential in information system. It can be best defined as the automated process of extracting useful knowledge and information including, patterns, associations, trees, changes, trends, anomalies and significant structures from large or complex data sets that are not classified[2]. Our main idea is that the hidden patterns, associations, classification and anomalies that are discovered by data mining techniques can help bridge this knowledge gap in higher learning institutions.[7] The knowledge discovered by data mining techniques would enable the learning institutions in making better decisions, having more advanced planning in directing students, predicting individual behaviors with higher accuracy, predicting the dropout rates and enabling the institution to allocate resources and staff more effectively.[4] It results in improving the quality, effectiveness and efficiency of the processes. The term data mining is often used to apply to the two separate processes of knowledge discovery and prediction. Knowledge discovery provides explicit information that has a readable form and can be understood by a person at user end[6]. Forecasting, or predictive modeling provides predictions of future events which can help in betterment and may be transparent and readable in some approaches and opaque in others such as neural networks.[7]This paper presents how various data mining techniques can be implemented in the field of higher education to discover some meaningful patterns or relations that can further improve the overall performance and quality of education and students 'respectively.

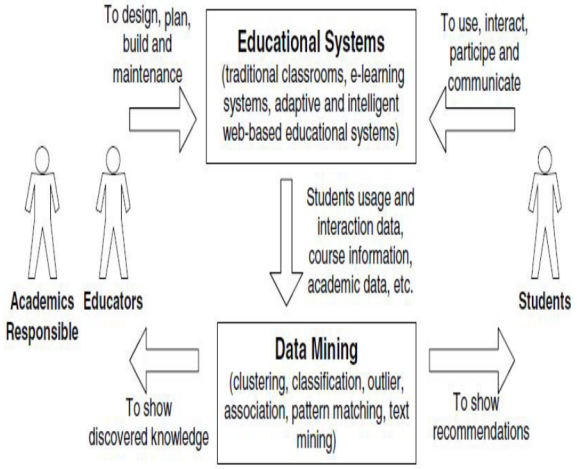


Fig 2.1. The cycle of applying data mining in educational systems

III. PROPOSED ANALYSIS GUIDELINE (DM-HEDU) FOR APPLICATION OF DATA MINING IN HIGHER LEARNING INSTITUTION

In this section, we propose a new analysis guideline to present a roadmap or the area of data mining application in higher learning institution [1]. Its adopted and primary name isDM-HEDU (Data Mining in Higher Education System)[1]. As today's higher learning institutions deal with powerful and strong business competitors in a highly competitive environment, they have to look for a new, faster and innovative solution to overcome the problems and achieve a high academic institutional standard. Therefore, this guideline may assist the institutions and organizations to identify the ways to improve their processes and help to take decisions [9]. In the previous literature studies we have not discovered complete guideline which gathers most of the possible processes to improve the level of higher education learning institution through data mining. Following table shows the attributes which can be used in higher education learning:

©IJRASET 2015: All Rights are Reserved

Volume 3 Issue IV, April 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

ID	Objective	Data Mining Method	Explicit Knowledge	Educational Process
1	Use of Data Mining in correct scores	Prediction	The patterns of previous student test score associated with their marks, attendance ,extra-curricular activities and so on.	Planning-Course assessment.
2	Creating meaningful learning outcome typologies	Cluster analysis	The patterns are generated based on previous student's learning outcome	Evaluation-Student assessment
3	Predicting students' overall performance	Classification	Classified pattern of previous students based on their performance throughout the year	Evaluation- Student assessment
4	Help for choosing the electives	Prediction	Based on the previous records and pattern generated, the appropriate elective is suggested for the student	Planning course assessment
5	Suggesting guidelines for teachers	Cluster analysis and Classification	Based on feedback given by the students various patterns are generated. Based on that patterns teaching pattern is observed and suggestions are given	Evaluation-Teacher assessment

Table 3.1: Summary analysis of previous study

A. ID 1: Uses of Data Mining in CRCT Scores

This study (Gabrilson, 2003) attempts to analyze the most effective factor in determining students' score in various subjects. It presents that the useful and meaningful discovered patterns targeting the various relationship of different types of variables are the major factors affecting the students test score. Using data mining prediction technique, these factors are thus successfully identified. From this case study the effectiveness of data mining in predicting the most effective and necessary factors in student test score can be concluded. It results in improving the evaluation and also helps in student assessment process. Improving this process has a direct impact in improving transition rate of a higher learning institution and thus decreases the drop-out rate.

B. ID 2: Creating Meaningful Learning Outcome Typologies

This case study aims at creating meaningful learning outcome typologies using data mining techniques. The main objective of obtaining typologies of students is to be able to improve students' performance through predefined clusters of behavior. These

©IJRASET 2015: All Rights are Reserved

clusters help higher education universities to better identify the requirements of each group and make better decision on how to behave with them in terms of educating, offering courses and curriculum, required time for teaching and so on.[8] It results in having more student satisfaction of their studies, course offering, and class's lectures. From this study we can conclude the effectiveness of data mining in developing typologies of students in higher educational domain. The result has a great impact in improving educational achievement of a higher education learning through improving the evaluation-student assessment process.

C. ID 3: Predicting a Student's Performance

This case study uses the data mining classification technique to predict the students final grades based on their web-use feature. By discovering the successful patterns of students in various categories, the institute can predict the final grade of each single student. Therefore it helps to identify students who are at risk early and allow the faculty to provide appropriate advice.

From this case study, we can conclude that data mining is effective in predicting student's performances in the educational domain. The result has deep impact in improving the transition rate, and the pro
cess indicator of a higher learning institute by improving the student assessment process to some extent.

D. ID 4: Help for choosing electives

This case study uses the data mining classification and prediction technique to predict the elective the student should choose based on their web-use feature. By discovering the successful patterns of students in various categories and by monitoring their academics the institute can predict almost the right and appropriate elective of each single student. Therefore it helps the students to choose the right elective and thus choose the right career path depending on the chosen elective.

From this case study the effectiveness of data mining in predicting the most appropriate elective can be concluded. It results in improving the evaluation and also helps in student assessment process. Improving this process has a direct impact in improving transition rate of a learning institution and thus decreases the drop-out rate and increase the overall performance of the student.

E .ID 5: Suggesting guidelines for teachers

This case study aims at creating meaningful learning outcome typologies using data mining techniques. The main objective of obtaining typologies of teachers is to be able to improve teachers' overall teaching pattern and performance through predefined clusters of behavior.[3] These clusters help higher education universities to better identify the pattern and the way of teaching of each and every teacher and make better decision on how to improve the teaching pattern incase they are lagging at some point. From this study we can conclude the effectiveness of data mining in improving the teaching pattern of the teachers. The result has a great impact in improving the Excellency of the teachers and thus leading to a better learning institution.

E. Advantage of Data mining in Academics

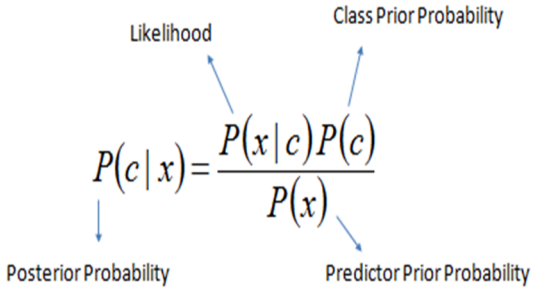
Data mining gives the answers of questions like:
Q: Who is the weak/strong student?
Q: Which is the interesting subject of the students?
Q: How can faculty help weak students?
Q: How the overall college result can be improved?
Q: What is the dropout rate and how it can be reduced?
Q: Is the teaching pattern satisfactory or need to be changed?
Q:Which is the most appropriate elective for the student?

IV. NAIVES BAYES ALGORITHM

Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifierassumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

©IJRASET 2015: All Rights are Reserved

Volume 3 Issue IV, April 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

A. P(c|x) is the posterior probability of class (target) given predictor (attribute).
B. P(c) is the prior probability of class.
C. P(x|c) is the likelihood which is the probability of predictor given class.
D. P(x) is the prior probability of predictor.

V. WEKA: A DATA MINING TOOL

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. Weka's main user interface is the Explorer, but essentially the same functionality can be accessed through the component-based Knowledge Flow.

A. Main features of Weka include
1) 49 data preprocessing tools
2) 76 classification/regression algorithms
3) 8 clustering algorithms
4) 15 attribute/subset evaluation + 10 search algorithms for feature selection
5) 3 algorithms for finding association rules
6) 3 graphical user interfaces
   -"The Explorer" (explanatory data analysis)
   -"The Experimenter" (experimental environment)
   -"The KnowledgeFlow" (new process model inspired interface)

B. Weka Application Interface
1) Explorer : preprocessing,attribute selection,learning,visualization
2) Experimenter : testing and evaluating machine learning algorithms
3) KnowledgeFlow : visual design of KDD process
4) Simple Command-line : A simple interface for typing commands

C. Weka data formats:
Attribute Relation File Format (ARFF) is the default file type for data analysis in weka but data can also be imported from various formats:
1) ARFF (Attribute Relation File Format) has 2 sections :
   -The Header information defines attribute name,type and relations
   -The Data section lists the data records

©IJRASET 2015: All Rights are Reserved

2) CSV : Comma Separated Values (text file)
3) Data can also be read from a database using ODBC connectivity

VI. DATA ANALYSIS AND INVESTIGATION

A. Domain Understanding

In this phase the higher education is wholly analyzed and the main data mining objectives are set and targeted accordingly.

B. Data Understanding

In this phase, the required raw data and attributes of students and faculties are collected based on the pre-defined objectives. According to our data mining goal, the raw data is related to:
1) Student demographic and academic knowledge.
2) Lecturer demographic and academic knowledge.
3) Course information and contents.
4) Semester status information and planning.
The data are then described and explored by (i) identifying pre-defined initial format of data, (ii) the meaning and description of individual attributes of student and faculty and (iii) determining the relation of attributes with each other[3]. The final part verifies the quality of data by determining the completeness and correctness of data.

C. Data Preparation

This phase of data mining is the final step of directly dealing with data.[8] The dataset produced in this section is used for modeling and the major analysis task. The importance of data preparation is to maximize visibility of the relationship that exists between input and output data sets, which is captured with a modeling tool.[5] Prepared data enables data mining technique to generate a better and efficient model.

VII. DATA MINING MODELING

The knowledge obtained from data mining techniques gives the managerial decision makers the useful information for taking making proper decisions [5]. The models are classified in two main categories: predictive models and descriptive models.

A. Descriptive model describes the data set in a concise and summarized manner and presents the interesting and important general properties of the data. It explains the extracted patterns in existing data, which may be used to guide managerial decisions.[4]
B. Predictive model predicts behavior based on previous data and uses data with known results to build a model that can be used in Future to explicitly predict values for different data (Two Crows Corporation, 1999).

A. Predictive Data Mining Models

Model A: Predicting Student Success Rate for Individual Student

This model is developed to predict the student success rate for individual students and to predict the student dropout rate[9]. Theexplicit knowledge discovered from this model can be used by student management systemto consult individual student based on his performance in successful course taking and choosing appropriate electives.

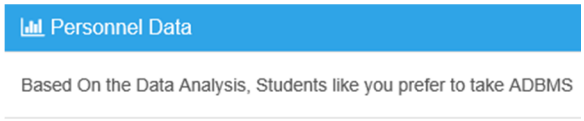


Fig 7.1. Elective Prediction result

Within this procedure, if the students are predicted to be unsuccessful, then they are provided with extra consultation and extra efforts are taken by faculty and the universities to help them to get improved in the course.

©IJRASET 2015: All Rights are Reserved

Volume 3 Issue IV, April 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

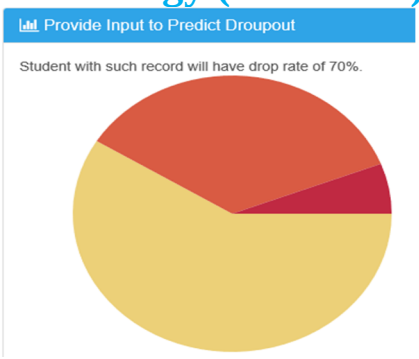


Fig 7.2. Dropout rate Prediction result

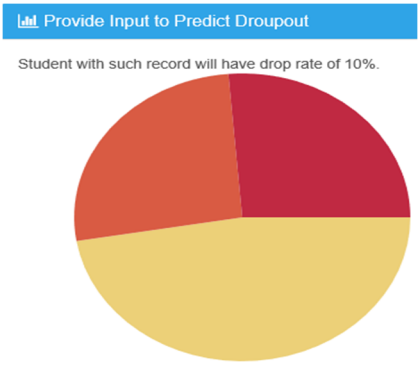


Fig 7.3. Dropout rate Prediction result

Model B: Model of Lecturer Course Assignment Policy Making

This model is developed to describe the characteristics pattern and way of teaching of lecturers who plan to take the course. [4]The knowledge discovered from this model can be used for general decision making at top-level management. It helps in knowing how the lecturer teaches.ie.is his teaching pattern fruitful or not. It assists in supporting the current managerial rules and regulations in lecturer course assignment policy making and it also helps to set new strategies and plan for managerial decision makers on those lecturers who plan to conduct the course.

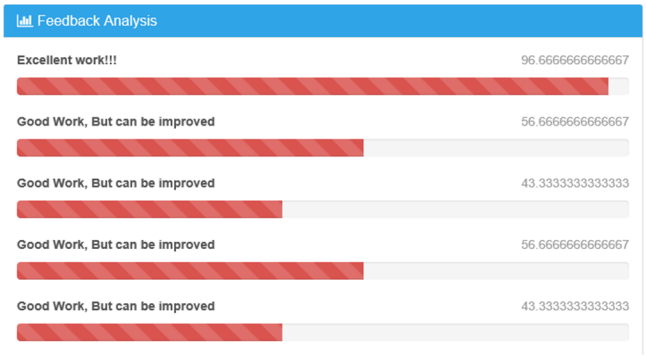


Fig 7.4. Feedback Analysis result

VIII. CONCLUSION AND FUTURE WORK

The current education system does not involve any prediction about fail or pass percentage of students based on their performance. The system doesn't deal with student dropouts. Since the proposed model identifies the weak and lagging students, the teachers can provide support and academic help for them. It also helps the teachers to act before a student drops or plan for recourse allocation

©IJRASET 2015: All Rights are Reserved

with confidence gained from knowing how many students are likely to pass or fail.Among several innovation in recent technology, data mining is making a great impact and comprehensive changes in the field of higher education. Such activities will definitely guide to better decision making procedures and will improve the quality of education.

As a further work, we would like to enhance other data mining processes in higher learning institution by referring to DM-HEDU analysis guideline. These processes are according to first class priorities of the institutions. Other work can be generating student and lecturer models for the other type of course offered in the institute. Since the application of data mining brings a lot of advantages in higher learning institution and even helps in improving the quality of students and education, it is recommended to apply these techniques in other academic institution like primary and secondary schools, language institutions, institutions for special students and private and government colleges especially in India.

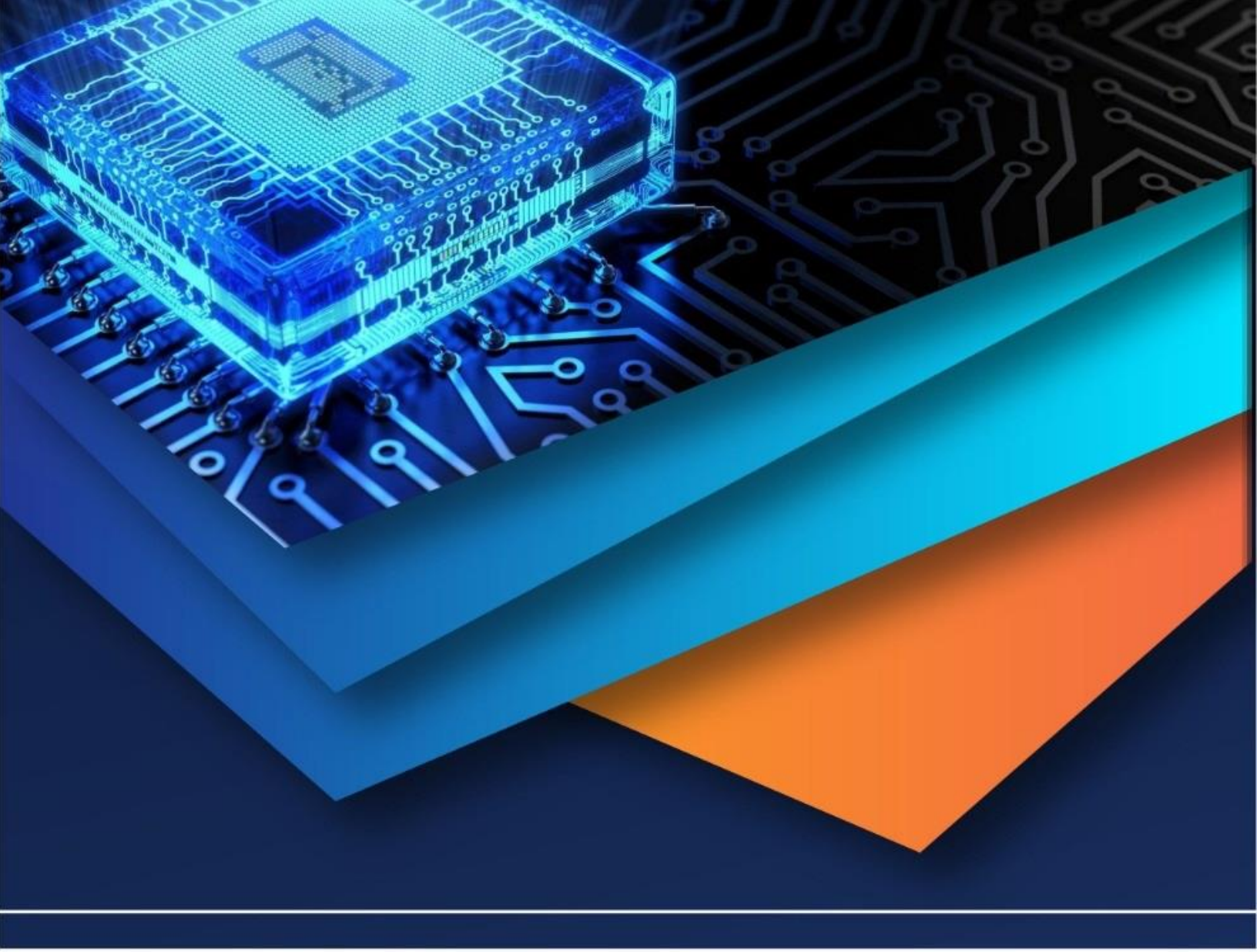
IX. ACKNOWLEDGEMENT

This work is carried under precious guidance of Prof. Priya Thakare, Dept. of Computer Engineering, SITS, Narhe. The authors are thankful for such helpful guidance and discussions at the early stage of the project.

REFERENCES

[1]Agrawal, R.T. and Imielinski, A.S. (1993). Mining association rule between sets of item in large database. In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D. C., 207–216.
[2]Cave, M., Kogan, M. and Hanney, S. (1990). The scope and effects of performance measurement in British higher education. In F.J.R.C. Dochy, M.S.R. Segers and W.H.F.W. Wijnen (Eds.), Management Information and Performance Indicators in Higher Education. Van Gorcum and Comp, B.V., Assen/Maastricht, 48–49.
[3]Chang, W.H.T. and Lee, Y.H. (2000). Telecommunications data mining for target marketing. Journal of Computers, **12**(4), 60–74.
[4]Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. and Wirth, R. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide.
[5]Chen, M.S., Han, J. and Yu, P.S. (1996). Data mining: an overview from a database perspective. IEEE Transaction on Knowledge and Data Engineering.
[6]Delavari, N. and Beikzadeh,M.R. (2004). A new analysis model for data mining processes in higher educational systems. In MMU International Symposium on Information and Communications Technologies 2004 in Conjunction with the 5th National Conference on Telecommunication Technology 2004. Putrajaya,Malaysia.
[7]Delavari, N., Beikzadeh, M.R. and Shirazi, M.R.A. (2004). A new model for using data mining in higher educational system. In 5th International Conference on Information Technology Based Higher Education and Training: ITEHT ’04. Istanbul, Turkey.
[8]Delavari, N., Beikzadeh, M.R. and Phon-Amnuaisuk, S. (2005). Application of enhanced analysis model for data mining processes in higher educational system. In 6th International Conference on Information Technology Based Higher Education and Training. Santo Domingo, Dominant Republic.
[9]Edelstein, H. (2000). Building profitable customer relationships with data mining. SPSS White Paper-Executive Briefing. Two Crows Corporation.
[10]Kononenko. Comparison of inductive and naive Bayesian learning approaches to automatic knowledge acquisition. Current Trends in Knowledge Acquisition, 1990.
[11]Pernkopf, F.: Bayesian network classifiers versus selective k-NN classifier. Pattern Recognition ,January 2005
[12]G. Holmes; A. Donkin and I.H. Witten (1994). "Weka: A machine learning workbench". Proc Second Australia and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia. http://www.cs.waikato.ac.nz/~ml/publications/1994/Holmes-ANZIIS-WEKA.pdf. Retrieved 2007-06-25.

©IJRASET 2015: 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)