iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** II **Month of publication:** February 2017

DOI: http://doi.org/10.22214/ijraset.2017.2059

www.ijraset.com

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

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

Data Mining in Higher Education Sector

Peeyush Vyas
Asst. Professor, CE/IT Department of Vadodara Institute of Engineering, Vadodara

Abstract: ***Data mining is the process of extracting hidden and useful information from a huge amount of data. It analyzes the data from different sources and convert them into significant information. Today, there is a big challenge for the institutions for predicting the career path related to higher education of the students. So, it would be a big task for the institutions to find the students who will be suitable for a particular course according to their expertise & area of interests and to find the students who will need more guidance and counselling in a specific area. Management of institution also need to have more information related to the results of the students, their success rate in the new courses offered etc. Here, the problem faced by students and higher education institutions for higher education will be discussed. So, 'Data Mining' is to solve all these types of problems. It is a better tools to analyze & predict result of the students and helps to institutions in the decision making more smartly and accurately. In this paper, there is a small try to make use of one of the data mining techniques-J48 algorithm to predict the result of the students along with Data Mining tool 'Weka' which will be used for the classification of the data.***
Keywords : Knowledge Discovery, Higher Education, Classification, J48 Algorithm, Weka.

I. INTRODUCTION

The overall goal of the data mining technique is to extract hidden information from a large data set and transform it into an understandable structure for further use. For the institutions or universities, Data Mining tools can predict future trend to make knowledge-driven decisions. So, these tools can save the time and can answer any kind of query in the less time as compare to the traditional procedures, helping in the decision support systems. The amazing predictive information and hidden patterns pertaining to academics can be found using Data mining tools. Using these tools, institution can predict with more accuracy that which student will be more suitable for a particular higher course or who will become graduate. This way, institution or university can use this information to pay more attention towards the weak students as well. To understand the concept of Data Mining let us first know some of the fundamental concept- Classification:

Classification is a data mining (machine learning) technique used to predict group membership for data instances. It identifies associations and clusters and separates subjects under study. Popular classification techniques include decision trees and neural networks. Classification models predict categorical class labels; and prediction models predict continuous valued functions. It is a supervised technique where we will design our model.\

Categorization

Categorization uses rule induction algorithms to handle categorical results, like "Persist" or "Dropout," and "Transfer" or "Stay."

Estimation

Estimation includes predictive functions and deals with continuous outcome variables, such as GPA and salary level.

Visualization

Visualization uses interactive graphs to show mathematically induced rules and scores. These graphs are more refined than any traditional bar chart or a pie chart. To show 3D geographic locations of mathematical coordinates, visualization is the bes technique.

So, higher education institutions can make use of classification for a comprehensive analysis of student characteristics, or use estimation to predict the likelihood of a variety of outcomes, such as transferability, persistence, retention, and course success. Other than these, various Data Mining tools and algorithms can be as follows –

Machine Learning

Computer Science, heuristics and Induction algorithm Neural Networks

Artificial Intelligence etc.

Biological models, psychology and engineering

Phases of Data Mining

Data mining is an iterative process that typically involves the following phases:

Problem definition Data exploration Data preparation Modeling

Volume 5 Issue II, February 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Fig. 1

A data mining mission begins with the understanding of the problem. Specialists of data mining, business experts, and domain experts work together to describe the project goals and also they find out the requirements from the business point of view. The various phases can be explained as below –

Problem Definition→In this project, our area or domain is data related to the field of academic such as records of students, results of the college, strength of the students year wise or department wise etc. The specialist members or experts can be considered as heads of the departments and principal of the college.

Data Exploration→In this phase, data are explored using the conventional tools like statistics.

Data Preparation→this phase allows to squeeze data many times. The main aim is simply to prepare data for modeling tool while selecting various databases, tuples and attributes without altering the basic meaning of the original data. If require, different types of mining functions can also be used.

Modeling→here, domain experts make some interaction with the data preparation phase to evaluate the model. If the proposed model does not meet the requirements and expectations and model is rebuilt while changing the suitable considerations until the desired model is made. We, finally, deploy the model, when we are satisfied.

A. Data collection & analysis tools

The following are the various types of tools required to perform the various tasks of the project like analyzing data, designing data, data implementation, developing software etc.

Weka Data Mining Tools, Tanagra Data Mining Tool, Web Miner, Ms-Access, MySql, SPSS etc.

B. Data mining experiment

Herein, some data is collected from one of the best colleges of Baroda for the research work. First of all, data are cleaned and integrated. The following table shows the various attributes with their respective values taken for the problem.

Table I: Selected Attributes

S.No.	Attribute	Description
1	Secondary_per	Low, Average, High
2	Sen_Sec_per	Low, Average, High
3	Gender	Male, Female
4	Attendance	Below, Good, Average
5	MidTerm1	A,B,C,D,F
6	MidTerm2	A,B,C,D,F
7	State	Gujarat, Rajasthan, Other
8	Result	Fail, R, Pass Low, Pass Average, Pass High

Fig. 2

Volume 5 Issue II, February 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

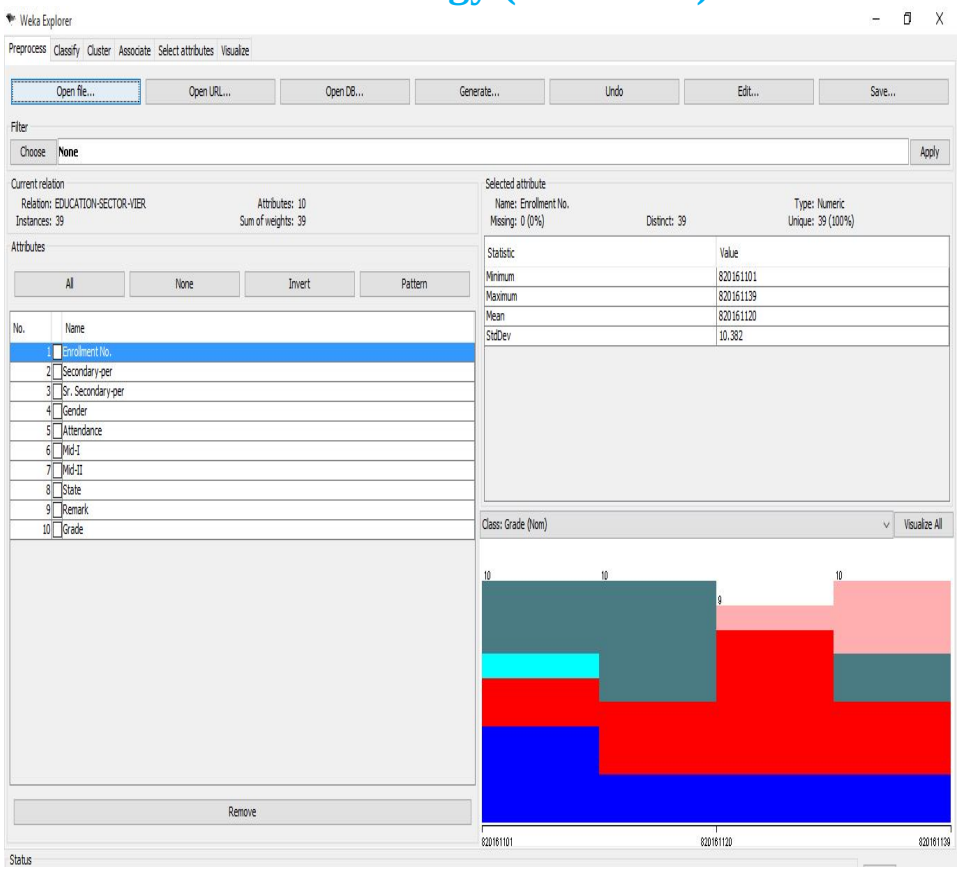


Table II: Sample Data

No.	1: Enrollment No. Numeric	2: Secondary-per Nominal	3: Sr. Secondary-per Nominal	4: Gender Nominal	5: Attendance Nominal	6: Mid-I Nominal	7: Mid-II Nominal	8: State Nominal	9: Remark Nominal	10: **Grade** Nominal
1	8.20161101E8	HIGH	HIGH	FEMALE	AVERAGE	A	A	GUJARAT	PASS	A+
2	8.20161102E8	LOW	LOW	FEMALE	AVERAGE	C	B	GUJARAT	FAIL	F
3	8.20161103E8	AVERAGE	AVERAGE	MALE	POOR	B	C	GUJARAT	PASS	B
4	8.20161104E8	AVERAGE	AVERAGE	MALE	GOOD	B	A	GUJARAT	PASS	A
5	8.20161105E8	HIGH	HIGH	MALE	GOOD	A	A	GUJARAT	PASS	A+
6	8.20161106E8	HIGH	HIGH	MALE	GOOD	A	A	GUJARAT	PASS	A+
7	8.20161107E8	HIGH	HIGH	MALE	GOOD	A	A	GUJARAT	PASS	A+
8	8.20161108E8	HIGH	HIGH	MALE	AVERAGE	A	B	GUJARAT	PASS	A
9	8.20161109E8	AVERAGE	AVERAGE	MALE	POOR	B	C	GUJARAT	FAIL	F
10	8.2016111E8	HIGH	HIGH	MALE	AVERAGE	A	B	GUJARAT	PASS	A
11	8.20161111E8	HIGH	HIGH	MALE	AVERAGE	A	B	GUJARAT	PASS	A
12	8.20161112E8	LOW	LOW	MALE	GOOD	C	B	GUJARAT	FAIL	F
13	8.20161113E8	LOW	LOW	MALE	POOR	C	B	GUJARAT	FAIL	F
14	8.20161114E8	LOW	LOW	MALE	POOR	C	C	GUJARAT	FAIL	F
15	8.20161115E8	HIGH	HIGH	FEMALE	AVERAGE	A	B	GUJARAT	PASS	A
16	8.20161116E8	HIGH	HIGH	FEMALE	AVERAGE	A	B	GUJARAT	PASS	A
17	8.20161117E8	HIGH	HIGH	FEMALE	POOR	A	A	GUJARAT	PASS	A+
18	8.20161118E8	HIGH	HIGH	FEMALE	AVERAGE	A	A	GUJARAT	PASS	A+
19	8.20161119E8	AVERAGE	AVERAGE	FEMALE	AVERAGE	B	A	GUJARAT	PASS	A
20	8.2016112E8	AVERAGE	AVERAGE	FEMALE	AVERAGE	B	A	GUJARAT	PASS	A
21	8.20161121E8	AVERAGE	AVERAGE	MALE	AVERAGE	B	B	GUJARAT	PASS	B+
22	8.20161122E8	LOW	LOW	MALE	AVERAGE	C	C	GUJARAT	FAIL	F
23	8.20161123E8	HIGH	HIGH	FEMALE	AVERAGE	C	B	GUJARAT	FAIL	F
24	8.20161124E8	HIGH	HIGH	FEMALE	GOOD	A	A	GUJARAT	PASS	A+
25	8.20161125E8	HIGH	HIGH	MALE	GOOD	A	A	GUJARAT	PASS	A+
26	8.20161126E8	LOW	LOW	MALE	POOR	C	C	GUJARAT	FAIL	F
27	8.20161127E8	LOW	LOW	FEMALE	POOR	C	C	GUJARAT	FAIL	F
28	8.20161128E8	LOW	LOW	FEMALE	POOR	C	B	GUJARAT	FAIL	F
29	8.20161129E8	LOW	LOW	MALE	POOR	C	B	GUJARAT	FAIL	F
30	8.2016113E8	LOW	LOW	MALE	POOR	C	C	GUJARAT	FAIL	F
31	8.20161131E8	AVERAGE	AVERAGE	FEMALE	AVERAGE	B	A	GUJARAT	PASS	A
32	8.20161132E8	AVERAGE	AVERAGE	FEMALE	AVERAGE	B	A	GUJARAT	PASS	A
33	8.20161133E8	AVERAGE	AVERAGE	FEMALE	AVERAGE	B	B	GUJARAT	PASS	B+
34	8.20161134E8	HIGH	HIGH	FEMALE	HIGH	A	A	GUJARAT	PASS	A+
35	8.20161135E8	LOW	LOW	MALE	POOR	C	C	GUJARAT	FAIL	F
36	8.20161136E8	AVERAGE	AVERAGE	MALE	AVERAGE	B	B	GUJARAT	PASS	B+
37	8.20161137E8	AVERAGE	AVERAGE	MALE	AVERAGE	B	B	GUJARAT	PASS	B+
38	8.20161138E8	HIGH	HIGH	FEMALE	HIGH	A	A	GUJARAT	PASS	A+

Fig. 3

Volume 5 Issue II, February 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

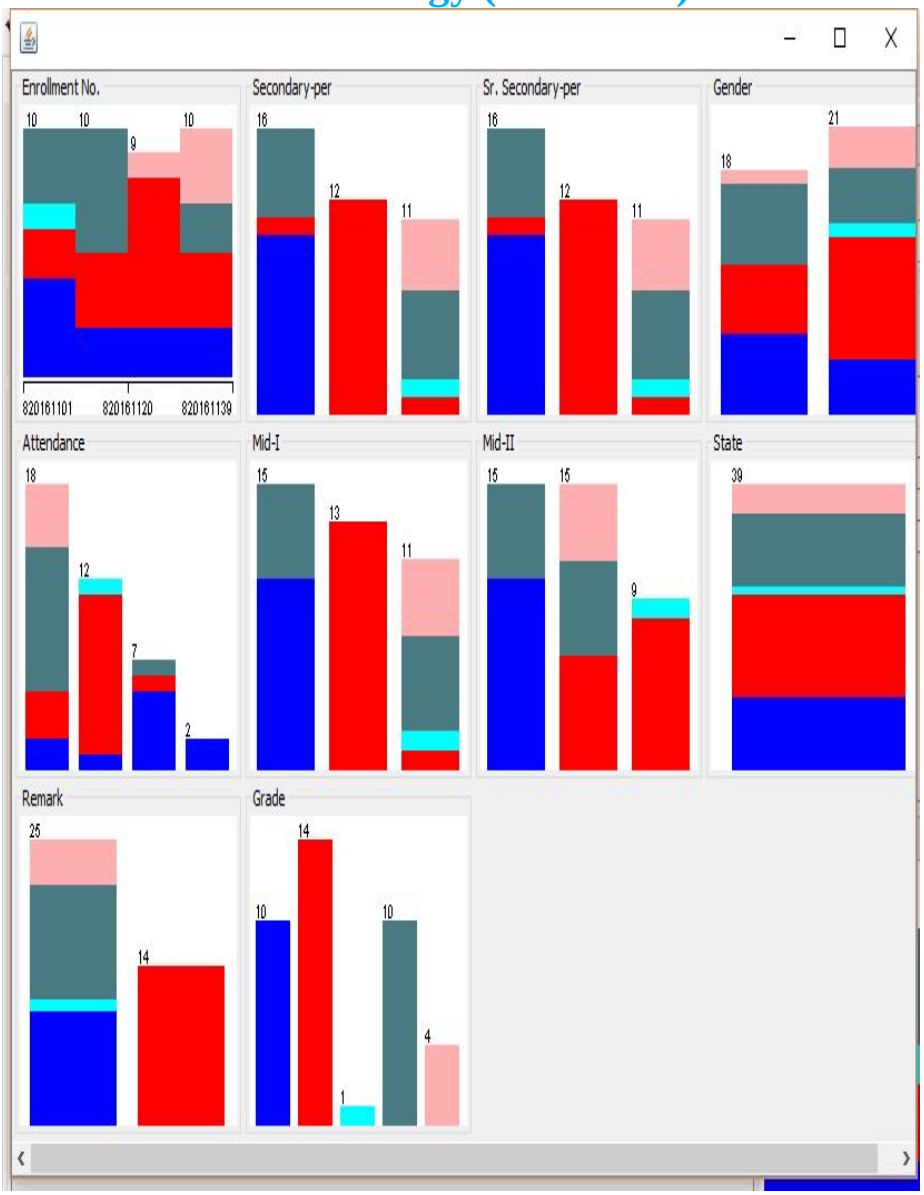
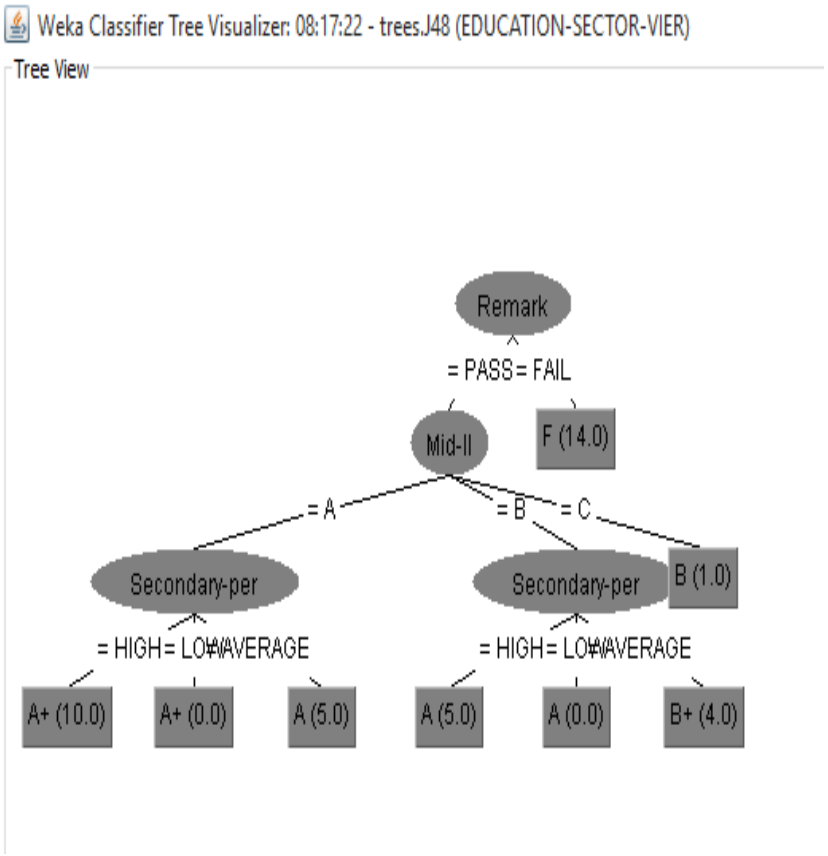


Fig. 4

Once data are integrated and cleaned, they are classified using data mining tool 'Weka'. For classification and predication J48 algorithm is used which is based on C4.5 algorithm from the machine learning. J48 algorithm is widely used algorithm of the Weka tool. It provides great strength between accuracy, speed and predictions of the results. It uses decision tree and using decision tree we can easily identify those students who are likely to be failed or we can predict the weak or brilliant students. On the basis of the current profile of the students, classification process can also be used to predict the chances of success or failure to pass the academic exams based on their profile. With the help of the decision tree we can easily find out those students who have maximum chances of failure or we can identify the weak students. With the help of this type of conclusion we can work hard on those students so that their performance and results can be improved. So, we can quote answers to the following questions with the help of data mining –

What are the interesting subjects of the students?
Who is the weak student How to improve result of the college?
Predicting the results?
How can weak students can get the help?

Volume 5 Issue II, February 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)



Statistical Results Produced By J48 Algorithm –

Fig. 5

```
Classifier output

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     EDUCATION-SECTOR-VIER
Instances:    39
Attributes:   10
              Enrollment No.
              Secondary-per
              Sr. Secondary-per
              Gender
              Attendance
              Mid-I
              Mid-II
              State
              Remark
              Grade
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

Remark = PASS
|   Mid-II = A
|   |   Secondary-per = HIGH: A+ (10.0)
|   |   Secondary-per = LOW: A+ (0.0)
|   |   Secondary-per = AVERAGE: A (5.0)
|   Mid-II = B
|   |   Secondary-per = HIGH: A (5.0)
|   |   Secondary-per = LOW: A (0.0)
|   |   Secondary-per = AVERAGE: B+ (4.0)
|   Mid-II = C: B (1.0)
Remark = FAIL: F (14.0)
```

Fig. 6

Volume 5 Issue II, February 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Classifier output

```
Number of Leaves  :     8

Size of the tree :      12


Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances          38               97.4359 %
Incorrectly Classified Instances         1                2.5641 %
Kappa statistic                          0.9645
Mean absolute error                      0.0103
Root mean squared error                  0.0838
Relative absolute error                  3.4607 %
Root relative squared error             21.8135 %
Coverage of cases (0.95 level)          97.4359 %
Mean rel. region size (0.95 level)      21.0256 %
Total Number of Instances               39

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1          0.034      0.909      1          0.952      1         A+
                 1          0          1          1          1          1         F
                 0          0          0          0          0          0.5       B
                 1          0          1          1          1          1         A
                 1          0          1          1          1          1         B+
Weighted Avg.    0.974      0.009      0.951      0.974      0.962      0.987
```

```
=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 10  0  0  0  0 |  a = A+
  0 14  0  0  0 |  b = F
  1  0  0  0  0 |  c = B
  0  0  0 10  0 |  d = A
  0  0  0  0  4 |  e = B+
```

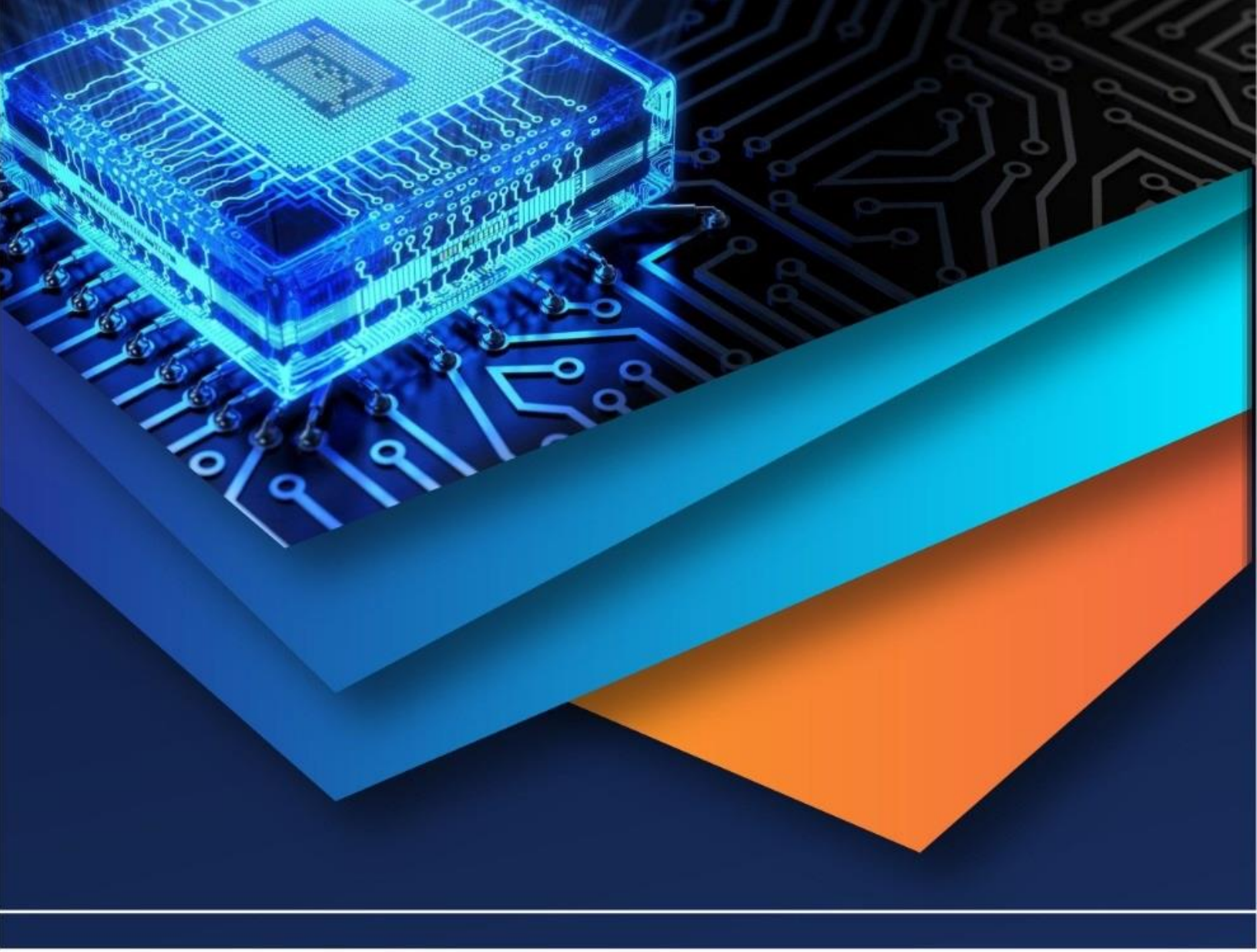
II. CONCLUSION

In the present scenario, there is no provision in the education system for predicting the result of the students on the basis of their performance. There is no computerized method which can predict and warn the students as well as faculty members about the low attendance or poor performance of the students. The functionalities used in this research paper not only help these areas but also it helps in selecting suitable higher course for the students on the basis of their area of interest and expertise. The students who are not counselled or guided properly are sometimes feel lost in the mass of the university or college. Simply, they are not able to go through their respective course and also they struggle to choose right career path. The proposed model discussed in this paper identifies such students and faculty members can better help in academics to them. The proposed system visualize data very nicely which can help students, college or university to take the significant decisions. Other than this, for future we can make the use of clustering techniques and can find particular domain and area of interest of students.

Volume 5 Issue II, February 2017
ISSN: 2321-9653

REFERENCES

[1] Hideko Kitahama, "Data Mining through Cluster Analysis Evaluation on Internationalization of Universities in Japan".
[2] Bruce L. Golden R. H. Smith School of Business University of Maryland College Park, MD 20742 "An Example of Visualization in Data Mining"
[3] Jing Luan, PhD Chief Planning and Research Officer, Cabrillo College Founder, Knowledge Discovery Laboratories "Data Mining Applications in Higher Education"
[4] Thulasi Kumarthulasi.kumar@uni.edu, University of Northern Iowa " Theoretical Basis for Data Mining Approach to Higher Education Research".
[5] N.V.Anand Kumar Research Scholar, Department of Computer Science and engineering Anna university, Chennai G.V.Uma Assistant professor, Department of Computer Science and Engineering Anna university, Chennai "Improving Academic Performance of Students by Applying Data Mining Technique".
[6] Brijesh Kumar Bhardwaj, Saurabh Pal "Data Mining: A prediction for performance improvement using classification" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 201
[7] S. Anupama Kumar and M. N. Vijayalakshmi "RELEVANCE OF DATA MINING TECHNIQUES IN EDIFICATION SECTOR" International Journal of Machine Learning and Computing, Vol. 3, No. 1, February 2013
[8] Mrs. Bharati M. Ramageri "DATA MINING TECHNIQUES AND APPLICATIONS" Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305.
[9] Ramageri, M. Bharti "Data Mining Techniques and App lications." Indian Journal of Computer sciences and Engineering, Vol.1 No.4, (Dec-Jan 2010- 2011): 301-305.
[10] Samira, AlEtia, Ipperciel, Donald and Hammad, Ahmed. "Implications and challenges to suing data mining in educational research in the Canadian context." Canadian Journal of Education, 35, 2 (2012): 101-119



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY