Journal of Theoretical and Applied Information Technology
31st December 2017. Vol.95. No 24

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# KNOWLEDGE DISCOVERY OF THE STUDENTS ACADEMIC PERFORMANCE IN HIGHER EDUCATION USING INTUITIONISTIC FUZZY BASED CLUSTERING

**T.PRABHA, Dr.D.SHANMUGA PRIYAA**

Research Scholar, Department of Computer Science, Karpagam Academy Of Higher Education, India
Professor, Department of Information Technology, Karpagam Academy Of Higher Education, India

**ABSTRACT**

This study observes the aspects associated with the evaluation of Students performance. To improve the performance of Students, their previous records have to be analyzed in order to determine their behavioral pattern in academic wise. This may help to assist the difficulty faced by the students to produce more marks in the semester exams and to enhance their co-curricular activities. This is done by clustering the students based on their performance. For performing the clustering process, this work utilized the fuzzy K-Mediods which takes the membership value of each student in account towards the particular cluster. The main contribution of this work is assisting to determine the level of performance of students and filtering the students who are in need of special attention by the staffs which results in improvising the quality of education.
**Keywords:** *Education mining, Student DataBase, Performance, Fuzzy K-Mediods, Academic*

## 1. INTRODUCTION

Digging of the hidden knowledge from a voluminous data to get useful information is the main objective of knowledge discovery [1]. Identifying useful pattern in a given data set is done using set of procedures in the means of algorithm is known as data mining. The popularity of data mining is flourished in the field of various applications. This paper adopts the concept of data mining in field of education which is now growing tremendously. This technique enhances the analyzing capability of the student's activities and infers the knowledge about their performance which helps in improvising them in terms of academic.

The concept of educational data mining [2] pushes the border of data mining to handle educational institutions data. Its scope is efficiently utilized to find the impact on the academic activities of the students it also encompasses the other area in educational process namely students admission coverage, relationship with alumni and pattern of selecting courses by the students [3]

The main focus of this research paper is to infer the interesting patterns of student's academic performance by exploring the hidden knowledge about the students and train them accordingly to produce good results. Hear by the technique of clustering the data set of the students for observing similarity pattern among them in order to evaluate their academic performance. The education data mining uses the following attributes from the student database of Karpagam University like Overall Semester Mark, Age, Day Scholar/hostler, Board Studied, Attendance, Co-curricular Activities (Paper Presentation, Seminar attendance) and Standing Arrears. The contribution in this paper is to utilize the clustering technique which aids in predicting the performance of students performance in the upcoming semester with their obtained grades in overall courses. As an outcome of this perception this research work digs the hidden knowledge of the student's academic performance and revealing this information give guidance to Students, Staffs and University's managements to take an appropriate action to improve their performance. In this paper, data is collected and analyzed from the students of Information Technology Department, at Karpagam University.

## 2. RELATED WORK

Mustafa Agaoglu [4] in their work designed a technique which mainly concentrates

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

only the performance of the students. According to the insight of the students evaluating the staffs performance by answering the questionnaire how they teach them, the efficiency of using teaching aids and the communication skill etc. In this education mining based classification technique is utilized to classify the performance of each staffs that handled the concerned courses. The results show that C4.5 classifier holds more accuracy than compared to other techniques used in their work. The outcome of this work reveals that most of the questions in questionnaire for evaluating the course are seem to be inappropriate.

Tripti Mishra et al [5] proposed a model which is highly focused on social integration of the students and their skills based on emotion and the academics. The Students third semester marks are considered for the evaluation and they used two different classification techniques namely j4.8 and random tree algorithm. The result shows that the random tree produces better result compared to j4.8.

Keno et al [6] developed a technique to determine the placement opportunity of the graduate students using nine different features. They collected the students database for the five years and in a random way they chosen the study of tracer form the placement office. The logistic regression based analysis is done on the selected dataset. Using this algorithm it is identified that gender, professional and core as potential features for selecting employability.

Bipin Bihari Jayasingh [7] identified the pattern of the behavior of the students by collecting the information from them by the means of questionnaire based on deductive learning. Using the relevancy of attributes and the rules with discriminant of class are identified using mining technique. The outcome is shown in charts and the output reveals that the learners of varying years have variant in the style of learning.

S. M. Merchán [8] developed a model for understanding the academic performance of the engineering students. It is done by constructing the predictive model which digs the hidden knowledge of the students by finding the characterization in their learning behavior.

Konstantina Chrysafiadi and Maria Virvou [9] proposed a new model based on online education for finding the programming knowledge of the students. The devised a module known as Fuzzy Knowledge State Definer for finding in dynamic manner and apprising the knowledge level of the students for whole concept of a particular domain expertize. They used Fuzzy based cognitive maps to represent dependency among the concepts of domains and exploring the students level of knowledge.

M. Mayilvaganan, D. Kalpanadevi [10] with the aid of education data mining they improved the performance of the prediction of students skill n academics they used weka tool kit to utilize the classifier algorithm for the selected dataset. They produced a comparative result of each algorithms performance in terms of accuracy.

Crist´obal Romero [11] they adapted EDM to find interesting patterns of subtends performance by studying various existing approaches on this filed. They prepared a survey on these relevant studies and provide useful information from their perspective of knowledge inferred from each model.

Pandey and Pal [12] gathered the dataset from different colleges under a university to determine the academic performance of the students using the bayes classifier based on the attributes category, language they studying , qualification background in order to find whether the new comers are not.

Galit [13] produce a study on result that will be obtained by the students based on their learning behavior and this helps to alert and warn the risk about their result in final exam and the way which they lack from other can be clearly known.

Bray [14], concentrates on private tutoring based performance analysis of the students in India compared to other countries. The study shows that the percentage of students prefer for private is comparatively high in India than other countries and it also noticed that performance of the students is enhancement and the intensity is based on the socio-economic criteria.

Pandey and Pal [15] they find the noticeable pattern of interest in student's performance using association rule based on in which language the subjects are taught.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Khan [16] they modeled a technique to analyze the performance of the senior secondary students based on the clustering scheme. From the result it is observed that girls with high socio-economic status had comparatively higher academic achievement in science stream and boys with low socio-economic status had fairly higher academic achievement in general.

Han and Kamber [17] explains the education data mining approach for analyzing the performance of the students using different direction and producing a distinct result on such identification.

Ayesha, Mustafa, Sattar and Khan [18] handled the learning activities of students by grouping their similarities using k-means clustering and predicts their performance and this aids in understanding the behavior of students by the instructor to teach them in a different aspect to produce good result and to enhance the students skill.

Bhardwaj and Pal [19] using 17 different features collected from the students the prediction of their performance in exams are shown by Bayesian classification

Al-Radaideh, et al [20] predicted the performance of the students in a particular course using the three different tree based models namely, id3, c4.5 and naïve bayes. The result shows that the Decision Tree algorithm had a improved forecasting than the other models.

## 3. PROPOSED METHODOLOGY

In the proposed system, the data related to all the students is being stored in a database. The performance of the student members is being displayed in the graphical form, so that the management and faculties can easily take decisions on improvising the performance of the students who lack in their studies with the help of this ED-mining technique. To conduct this analyze, the proposed work used questionnaire to collect the real data which describing the relationships between the behaviour of the students (psychometric factors) and their final academic performance, and contribution to a better quality education.

In Preprocessing Phase the action of identifying incomplete or incorrect data in a student's database. So it is an essential part to convert a raw dataset to complete dataset making it suitable for further analysis. Next step is to transform the data into a common format for processing. Data reduction, dimensionality reduction and the data conversion method can be used to diminish the number of conceivable values being measured. Clustering of Student database is an identification of similar classes of objects. It identifies dense and scarce regions in data space and can find out the global scattering design and correlations amongst the data attributes. Fuzzy K-mediods is to find out the prominent clusters of students. The discovered knowledge will report the Student's performance, useful suggestion to faculty and constructive recommendation to higher authorities of the University.

### Dataset Description

The data set used is obtained from B.Sc., Department of Information Technology(IT) 2015 to 2018 batch, Karpagam University, Coimbatore. Totally 58 students' records are collected and preprocessed to convert into a meaningful data. The collected attribute details are as follows:

SI.no, Register Number, Name (as per SSLC or HSC) , Date Of Birth, Parent's Name (Father), Parent's Name (Mother), Parent's Mobile (Father), Parent's Mobile (Mother), Student Mobile no, Address (Complete Permanent Address), Native State, Hostler or Dayscholar, Email, $10^{th}$ Std school (Complete Address), Board, Year of Passing in $10^{th\ std}$ , Maths in $12^{th\ std}$, $12^{th}$std.Mark % , School (Complete Address), Board, Year of Passing in $12^{th\ std}$ , Degree, Branch, Sem.1, Sem.2, Sem.3, Sem.4, CGPA upto 3rd Sem, Aggregate Percentage, total no.of arrears, No of arrears cleared, current standing arrears, break in study (yes / no) - if yes mention the number of years, training attended.

***Calculation***

Sum of the product of the GP by
the corresponding credits of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**GPA** = courses offered in that Semester / Sum of the credits of the courses of that Semester.

i.e. $\mathbf{GPA} = \frac{\sum_i CiGPi}{\sum_i Ci}$

**CGPA** = Sum of the product of the GPs by the corresponding credits of the courses offered for the entire programme / Sum of the credits of the courses Of the entire programme

i.e. $\mathbf{CGPA} = \frac{\sum_n \sum_i CniGPni}{\sum_n \sum_i Cni}$

Where,

$C_i$ = credit fixed per course 'i' in any semester

$GP_i$ = Grade point obtained for the course 'i' in any semester

'n' = refers to the Semester in which such courses are credited.

**Note: RA** grade will be excluded for calculating **GPA** and **CGPA.**

## K-mediods

The *k*-medoids algorithm is more relative to the two popular clustering techniques namely kmeans and medoid shift algorithm. Their function is to split the dataset and frame groups based on the similarity measure. Their aim is to reduce the distance between points within a particular cluster to the centroid point of the cluster. The unique feature of k-mediods is that they assume center as data points and uses Manhattan normalization to compute the distance between two data points under consideration for grouping. *k*-medoid is a conventional subdividing method of clustering that groups the data set of *n* objects into *k* clusters known in advance. While comparing to k-means, k-mediods are more vigorous to noise and outliers because it minimize them by adapting a sum of pair wise variations as an alternative of a sum of squared Euclidean distances used in k-means. A medoid can be well-defined as the object of a cluster whose typical dissimilarity to all the objects in the cluster is negligible. i.e. it is utmost centrally positioned point in the cluster.

## Fuzzy K-mediods

In Fuzzy K-mediods clustering algorithms it consist of two major processing phase. First, revealing a suitable function to discover every instances membership degree of all clusters. Second, attain a method that computes the cluster centers. Naturally the subsequent objective function is used as the membership degree calculating function.

$$P(Z,X) = \sum_{i=1}^{n}\sum_{j=1}^{k} \mu_{ij} r(x_j, z_i)$$

Where
$\mu_{ij}$ represents the degree associated with membership of the $i^{th}$ object $x_i$ to the $j^{th}$ cluster $z_j$, where Z contains the cluster centers, and $r(x_i\ z_j)$ is a difference measure between the $j^{th}$ cluster center and the $i^{th}$ object. Minkowski distance is used for dissimilarity measure between $x_i$ and $z_j$

Euclidian distance is applied for dissimilarity measure which is denotes as follows:

$$R(x_j,Z_i) = \sqrt{\sum_{l=1}^{m}(Z_{jl} - x_{jl})^2}$$

1. Select the initial medoids
2. Define the new medoid of each cluster to update medoids
3. Allocate each instance of object to the nearest medoid

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. Calculate sum of distance from all data objects to their medoids
5. Recap step 2 until the sum remains persistent

## 4. EXPERIMENTAL RESULT

The experimental Results were conducted using matlab and weka toolkit on the B.Sc IT students of Karpagam University Dataset. The performance of the proposed work is analyzed based on the various dimensions and the result shows that the fuzzy k-mediods based approach outperforms the existing k-means and k=Mediods.

### Evaluation Metric

For the comparison result three parameters are used and they are accuracy, precision and recall and their calculations are as follows

- Accuracy = (TP + TN) / (TP + TN + FP + FN) = #correct / #all_instances
- Precision = TP / (TP + FP) = #correct_positive / #classified_as_positive
- Recall = TP/ (TP + FP) = # correct_positive/#classified as correct_positive

The table shows the modeling output of the fuzzy k-mediods technique based on the performance of the students in overall semester. The cluster is done based on the six categories namely Excellent, Very good, Good, Above Average, Average and absent.

=== Model and evaluation on training set ===

| Cluster No | No. of Instances | Percentage |
|---|---|---|
| 0 | 14 | 24% |
| 1 | 4 | 7% |
| 2 | 23 | 40% |
| 3 | 1 | 2% |
| 4 | 3 | 5% |
| 5 | 13 | 22% |

The clustering is done of the overall performance of the students and the percentage of cluster 0 is 24 , cluster 1 is 7, cluster 2 is 40, cluster 3 is 2, cluster 4 is 5 and cluster 5 is 22.

## 5. CONCLUSION

In this paper, the performance of three different clustering algorithms namely: k-means, k-Mediods and proposed Fuzzy K-Mediods works on student performance data which are explored with a view to reveale accurate each algorithm could perform in grouping the dataset. The Result shows that the proposed Fuzzy K-Mediods performs better than other two approaches used in this paper. From the outcome of this work, most of the students belong to the state Board, Day Scholar and involvement in extracurricular activities is greatly high. Most of the students performance is good in their academic terms but when there is a lapse in attendance they failed to attend their semester marks. The fuzzy k-means appear to be the fastest among the two algorithms and is known for its excellent performance on large data, the algorithm however requires that, the value of k be continuously varied to get a cluster of good quality.

## REFERENCES

[1] Heikki, Mannila, .Data mining: machine learning, statistics, and databases., IEEE, 1996.
[2] Toon calders SIGKDD Explorations .Introduction to the Special Section on Educational Data Mining. Volume 13, Issue 2.
[3] Richard A. Huebner, “A survey of educational data- mining research”. Research in Higher Education Journal.
[4] Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," IEEE Access , Volume : 4 , 2016.
[5] Tripti Mishra,Dr. Dharminder Kumar,Dr. Sangeeta Gupta,"Mining Students’ Data for Performance Predict ion," in fourth International Conference on Advanced Computing & Communication Technologies,2014.
[6] Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat," Predicting IT Employability Using Data Mining Techniques," in third International Conference on Digital Information

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Processing, Data Mining, and Wireless Communications (DIPDMWC), 2016.

[7] Bipin Bihari Jayasingh,"A Data Mining Approach to Inquiry Based Inductive Learning Practice In Engineering Education," in IEEE 6th International Conference on Advanced Computing, 2016

[8] S. M. Merchán,"Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic," IEEE Latin America Transactions, vol. 14, no. 6, June 2016.

[9] Konstantina Chrysafiadi and Maria Virvou," Fuzzy Logic for adaptive instruction in an e-learning environment for computer programming," IEEE Transactions on Fuzzy Systems ,Volume: 23, Issue: 1, Feb. 2015.

[10] M. Mayilvaganan,D. Kalpanadevi ," Comparison of Classification Techniques for predicting the performance of Students Academic Environment," in International Conference on Communication and Network Technologies (ICCNT), 2014.

[11] Cristóbal Romero," Educational Data Mining: A Review of the State of the Art," IEEE Transactions On Systems, Man, And Cybernetics — Part C: Applications And Reviews, Vol. 40 , No. 6, November 2010

[12] U . K. Pandey, and S. Pal, .Data Mining: A prediction of performer or underperformer using classification., (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011.

[13] Galit.et.al, .Examining online learning processes based on log files analysis: a case study. Research, Reflection and Innovations in Integrating ICT in Education 2007.

[14] M. Bray, .The shadow education system: private tutoring and its implications for planners., (2nd ed.), UNESCO, PARIS, France, 2007.

[15] U. K. Pandey, and S. Pal, .A Data mining view on class room teaching language., (IJCSI) International Journal of Computer Science Issue, Vol. 8, Issue 2, pp. 277-282, ISSN:1694-0814, 2011.

[16] Z. N. Khan, .Scholastic achievement of higher secondary students in science stream., Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005.

[17] J. Han and M. Kamber, .Data Mining: Concepts and Techniques,. Morgan Kaufmann, 2000.

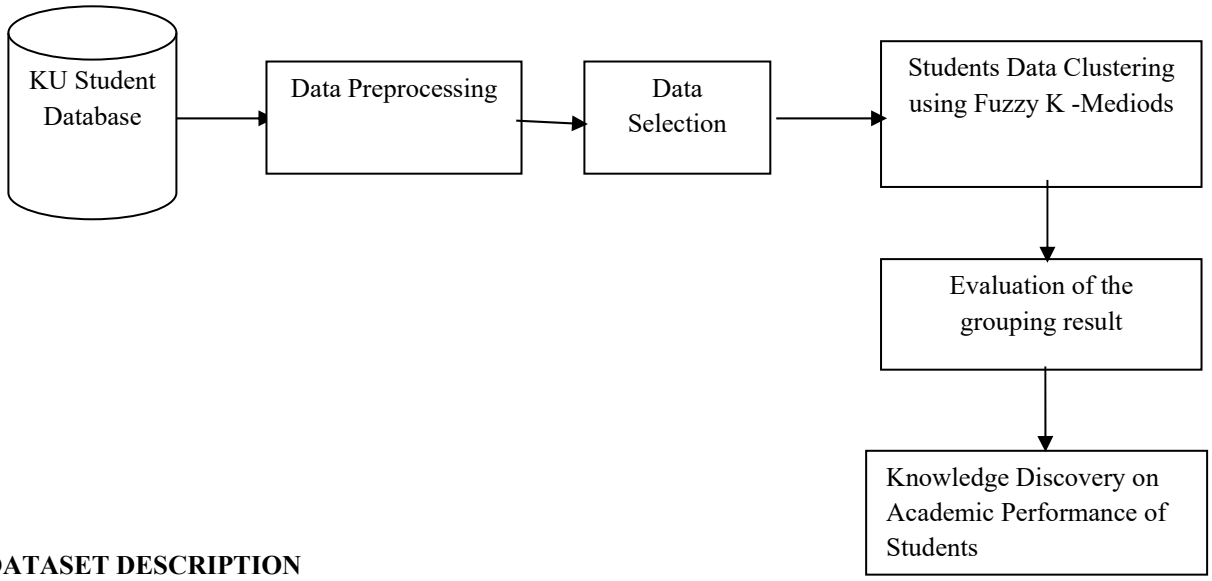
[18] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, .Data mining model for higher education system., Europen Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010.

[19] B.K. Bharadwaj and S. Pal. .Data Mining: A prediction for performance improvement using classification., International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[20] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, .Mining student data using decision trees., International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

## PROPOSED METHODOLOGY




## DATASET DESCRIPTION

| Fields | Field Description | Field possible values |
|---|---|---|
| AGE | Age of the student | Numeric (Calculated from the Date Of Birth) |
| BRD | Board of Studies | Alphabet |
| ESP | End Semester Performance | 91-100 (Outstanding) / O<br>81-90 (Excellent)/A+<br>71-80 (very Good) A<br>66-70 (Good) B+<br>61-65 (Above Average) B<br>55-60 (Average) C<br>50-54 (Pass) P<br><50 (Reappearance) RP<br>- (Absent) AB/AAA |
| CC | Co-curricular Activities/ Extracurricular Activities | Paper presentation, technical events, quiz, debugging / Sports, NSS, NCC |
| HS/DS | Hostler/Day scholar | {hostler, day scholar} |
| ATT | Attendance | <75 = 0 (poor)<br>75-80 = 2 (Average)<br>81-85 = 3 (Average)<br>86-90 =4 (Good)<br>91-100 =5 (Good) |
| AP | Aggregate Percentage | In percentage |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | |
|---|---|---|
| CSA | Current Standing Arrears | Numeric |
| CGPA | Cumulative Grade Point Average | Numeric |
| GRD | Grade | {O, A+, A,B+,B,C,D,RA,AA} |
| PRF | Overall Performance result | {Outstanding, Excellent, Very Good, Good, Above Average, Average, Pass, Reappearance, Absent} |

## RESULTS AND DISCUSSIONS

| S.no | Name | Age | BRD | ESP | CC | Hostler / Day scholar | ATT | CG PA | Grade | Aggregate Percentage | Current Standing Arrears | Performance |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | ABDUL HASITH M | 18 | State | A | A | DAYSCHOLAR | G | 5.88 | C | 58.8 | 0 | Average |
| 2 | ABIRAMI K | 20 | State | G | G | HOSTLER | G | 8.29 | A | 82.9 | 0 | Excellent |
| 3 | AISHWARYA A | 18 | State | G | A | HOSTLER | G | 7.71 | B+ | 77.1 | 0 | Good |
| 4 | AKASH G | 19 | State | A | P | DAYSCHOLAR | G | 5.85 | C | 58.5 | 0 | Average |
| 5 | ARAVINDHAN M | 19 | State | A | P | DAYSCHOLAR | G | 5.78 | C | 57.8 | 0 | Average |
| 6 | ARSHADH K S | 18 | State | p | P | DAYSCHOLAR | A | - | AB | - | 1 | Absent |
| 7 | BHARATHI PRIYA N C | 19 | State | G | G | DAYSCHOLAR | G | 7.12 | B+ | 71.2 | 0 | Good |
| 8 | BHARATHKUMAR U | 18 | State | P | P | DAYSCHOLAR | A | - | AB | - | 1 | Absent |
| 9 | BHUVANESHWARI P | 21 | State | G | G | DAYSCHOLAR | G | 8.33 | A | 83.3 | 0 | Very Good |
| 10 | DHANUSHKUMAR R | 19 | State | P | P | DAYSCHOLAR | G | - | AB | - | 2 | Absent |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | DHEEPIKA THARANI K | 19 | State | G | A | DAYSCHOLAR | G | 7.01 | B+ | 70.1 | 0 | Good |
| 12 | DIVYA S | 19 | State | G | A | DAYSCHOLAR | G | 7.33 | B+ | 73.3 | 0 | Good |
| 13 | FEMINA S | 19 | State | G | G | DAYSCHOLAR | G | 8.59 | A | 85.9 | 0 | Very Good |
| 14 | GAYATHRI A | 19 | State | G | P | DAYSCHOLAR | G | 7.1 | B+ | 71 | 0 | Good |
| 15 | GOKUL NATH P | 18 | State | P | P | DAYSCHOLAR | A | - | AB | - | 3 | Absent |
| 16 | HARSINI A | 19 | State | G | A | DAYSCHOLAR | G | 7.19 | B+ | 71.9 | 0 | Good |
| 17 | INDHUMATHI D | 19 | State | A | A | DAYSCHOLAR | G | 6.01 | B | 60.1 | 0 | Above Average |
| 18 | JAGADESH G | 21 | State | P | P | DAYSCHOLAR | A | - | AB | - | 1 | Absent |
| 19 | JAYALAKSHMI S | 18 | State | G | G | DAYSCHOLAR | G | 8.19 | A | 81.9 | 0 | Very Good |
| 20 | JEYA VINISTA J | 19 | State | G | G | DAYSCHOLAR | G | 8.32 | A | 83.2 | 0 | Very Good |
| 21 | KALPANA R | 18 | State | G | G | HOSTLER | G | 8.08 | A | 80.8 | 0 | Very Good |
| 22 | KARTHICK J | 19 | State | A | P | DAYSCHOLAR | G | 6.88 | B | 68.8 | 0 | Above Average |
| 23 | KARTHIK PRABU S T | 18 | State | A | P | DAYSCHOLAR | A | 5.3 | C | 53 | 0 | Average |
| 24 | KIRAN G | 19 | State | A | A | DAYSCHOLAR | A | 6.38 | B | 63.8 | 0 | Above Average |
| 25 | LAVANYA S | 19 | State | A | G | DAYSCHOLAR | G | 7.34 | B+ | 73.4 | 0 | Good |
| 26 | MANI MEGALAI K | 18 | State | G | A | DAYSCHOLAR | G | 7.79 | B+ | 77.9 | 0 | Good |
| 27 | MASANAHARI M | 18 | State | A | A | DAYSCHOLAR | G | 5.58 | C | 55.8 | 0 | Average |
| 28 | MUBEENA BANU. | 19 | State | A | G | DAYSCHOLAR | G | 7.29 | B+ | 72.9 | 0 | Good |
| 29 | MURALI P | 19 | State | P | P | DAYSCHOLAR | A | - | AB | - | 1 | Absent |
| 30 | NAGARAJ M | 20 | State | P | P | DAYSCHOLAR | G | - | AB | - | 1 | Absent |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 31 | NISAJ T L | 19 | State | A | P | DAYSCHOLAR | G | 6.27 | B | 62.7 | 0 | Above Average |
| 32 | NIVETHA G | 18 | State | G | G | DAYSCHOLAR | G | 8.45 | A | 84.5 | 0 | Very Good |
| 33 | NIVETHA P | 20 | State | G | G | DAYSCHOLAR | G | 8.11 | A | 81.1 | 0 | Very Good |
| 34 | NIVETHA K | 20 | State | G | G | DAYSCHOLAR | G | 7.51 | B+ | 75.1 | 0 | Good |
| 35 | PREETHA R | 19 | State | G | G | DAYSCHOLAR | G | 8.58 | A | 85.8 | 0 | Very Good |
| 36 | PRIYADHARSHINI S | 19 | State | G | G | DAYSCHOLAR | G | 9.356 | A+ | 93.56 | 0 | Excellent |
| 37 | PRIYADHARSHINI S | 19 | State | G | G | DAYSCHOLAR | G | 8.64 | A | 86.4 | 0 | Very Good |
| 38 | RAJALAKSHMI S | 19 | State | G | G | DAYSCHOLAR | G | 7.85 | B+ | 78.5 | 0 | Good |
| 39 | RENIN B | 19 | State | P | P | DAYSCHOLAR | A | - | AB | - | 4 | Absent |
| 40 | SABARINATH R | 19 | State | A | A | DAYSCHOLAR | A | 5.32 | C | 53.2 | 0 | Average |
| 41 | SAMEEL MURATH M A | 19 | State | G | A | DAYSCHOLAR | G | 6.77 | B | 67.7 | 0 | Above Average |
| 42 | SANDEEP R | 19 | State | G | A | DAYSCHOLAR | G | 7.04 | B+ | 70.4 | 0 | Good |
| 43 | SANGEETHA N | 20 | State | G | A | DAYSCHOLAR | G | 6.77 | B | 67.7 | 0 | Above Average |
| 44 | SANGEETHA T | 19 | State | G | G | DAYSCHOLAR | G | 7.44 | B+ | 74.4 | 0 | Good |
| 45 | SANTHAKUMAR R | 19 | State | A | G | DAYSCHOLAR | G | 5.93 | C | 59.3 | 0 | Average |
| 46 | SANTHANALAKSHMI S | 18 | State | G | A | DAYSCHOLAR | G | 6.86 | B | 68.6 | 0 | Above Average |
| 47 | SARIKA B | 19 | CBSE | G | A | DAYSCHOLAR | G | 7.15 | B+ | 71.5 | 0 | Good |
| 48 | SASIKUMAR S | 19 | State | A | P | DAYSCHOLAR | G | 6 | B | 60 | 0 | Above Average |
| 49 | SASIKUMAR B | 18 | State | A | P | DAYSCHOLAR | A | 5.92 | C | 59.2 | 0 | Average |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | SHIHAN HUSSAIN S | 20 | State | P | P | DAYSCHOLAR | A | - | AB | - | 2 | Absent |
| 51 | SHRUMOD R | 20 | State | P | P | DAYSCHOLAR | A | - | AB | - | 1 | Absent |
| 52 | SIVAPRAKASH S | 18 | State | P | P | DAYSCHOLAR | A | - | AB | - | 3 | Absent |
| 53 | SUBASHINI K | 19 | State | G | A | DAYSCHOLAR | G | 7.71 | B+ | 77.1 | 0 | Good |
| 54 | SUDHAKAR M | 19 | State | A | G | DAYSCHOLAR | G | 5.52 | C | 55.2 | 1 | Average |
| 55 | TAMILARASAN S | 19 | State | A | G | DAYSCHOLAR | A | 6.73 | B | 67.3 | 0 | Above Average |
| 56 | VISHNU BHARATH M | 21 | CBSE | P | P | DAYSCHOLAR | G | - | AB | - | 1 | Absent |
| 57 | YASER ARAFATH M | 21 | State | A | G | DAYSCHOLAR | A | 5.27 | C | 52.7 | 2 | Average |
| 58 | YOKESH K | 18 | State | P | P | DAYSCHOLAR | P | - | AB | - | 0 | Absent |

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**Confusion Matrix for Fuzzy K-Mediods with Performance as Class attribute**

| 0 | 1 | 2 | 3 | 4 | 5 | <-- assigned to cluster |
|---|---|---|---|---|---|---|
| 5 | 2 | 0 | 0 | 3 | 0 | **Average** |
| 0 | 0 | 1 | 1 | 0 | 0 | **Excellent** |
| 0 | 2 | 13 | 0 | 0 | 0 | **Good** |
| 0 | 0 | 0 | 0 | 0 | 13 | **Absent** |
| 0 | 0 | 9 | 0 | 0 | 0 | **Very Good** |
| 9 | 0 | 0 | 0 | 0 | 0 | **Above Average** |

Where,

Cluster 0 <-- Above Average
Cluster 1 <-- No class
Cluster 2 <-- Good
Cluster 3 <-- Excellent
Cluster 4 <-- Average
Cluster 5 <-- Absent

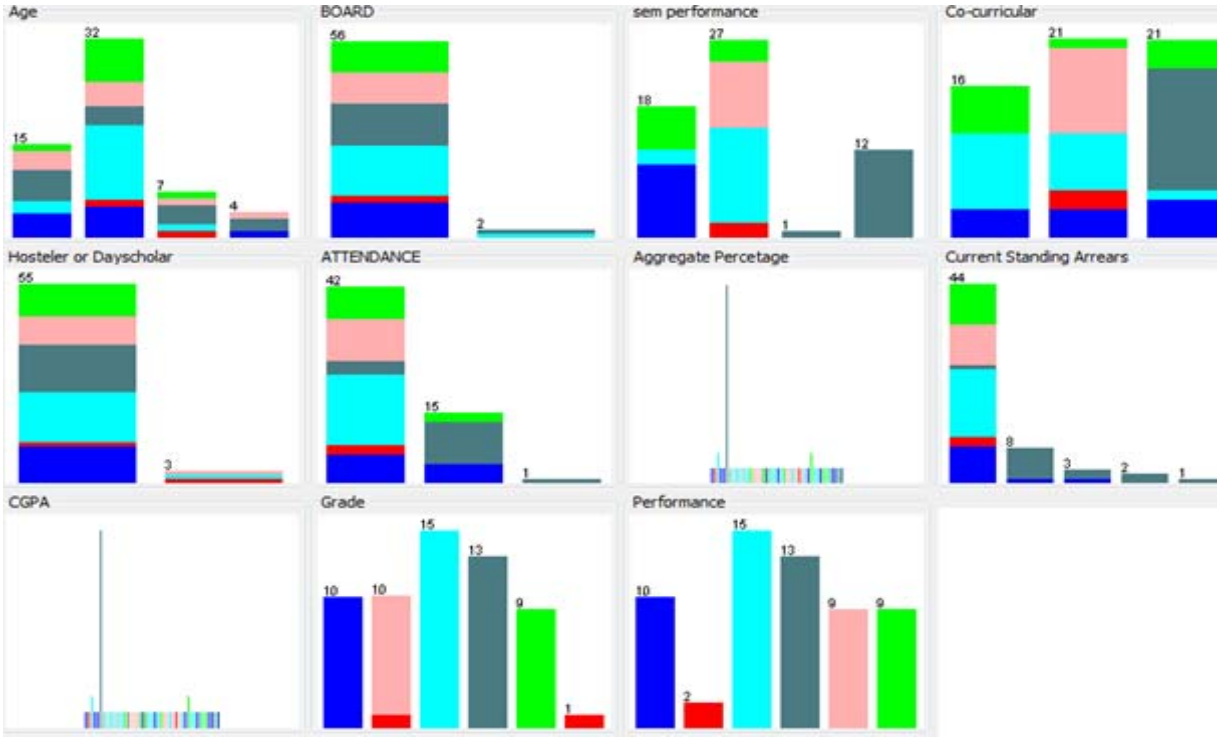
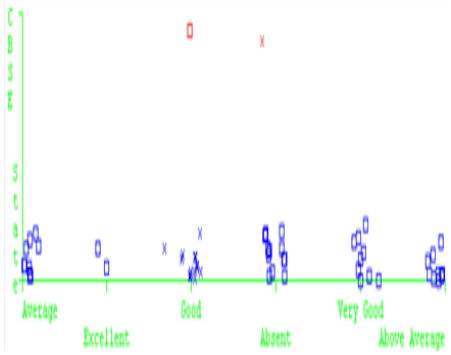
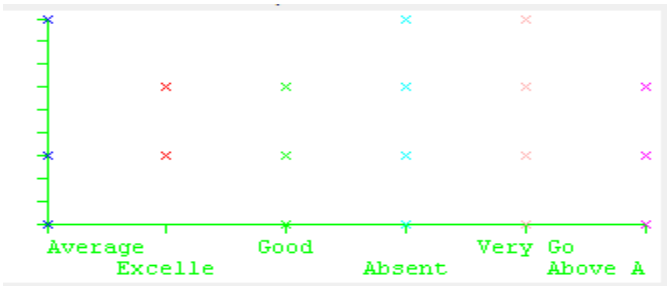



*Figure: Range of distribution of dataset on Student database*

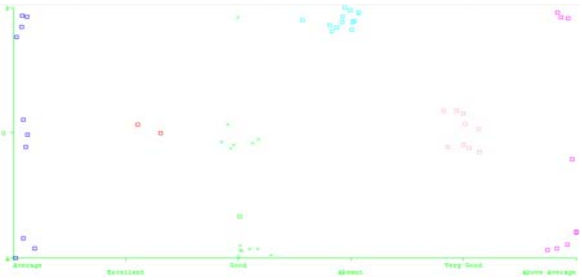
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



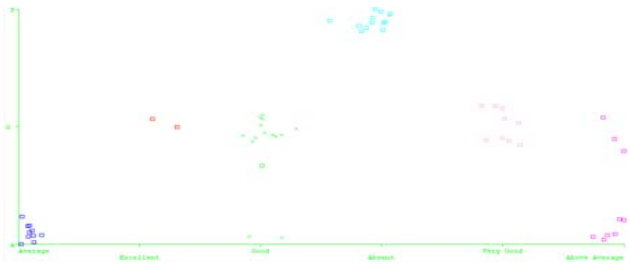

*a) Board Vs Performance*



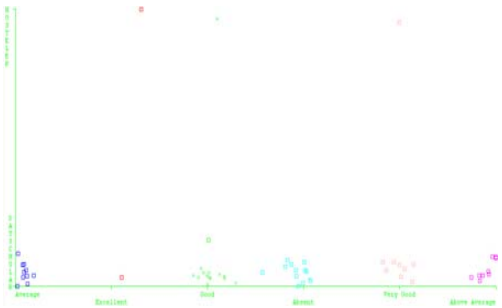

*b) Age Vs Performance*



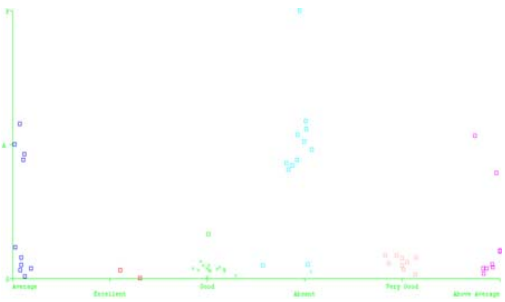
*c) End Semester Vs Performance*



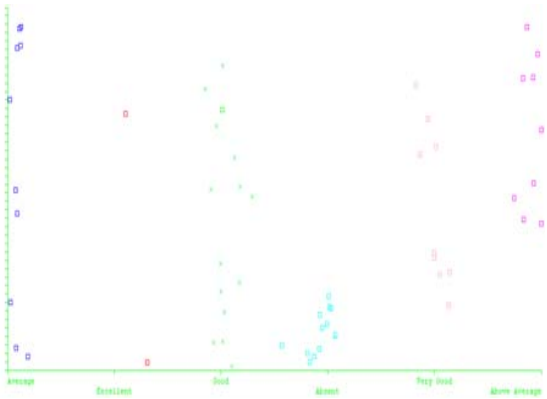
*d) Co-curricular Activities Vs Performance*



*e) Hostler/Day Scholar Vs Performance*

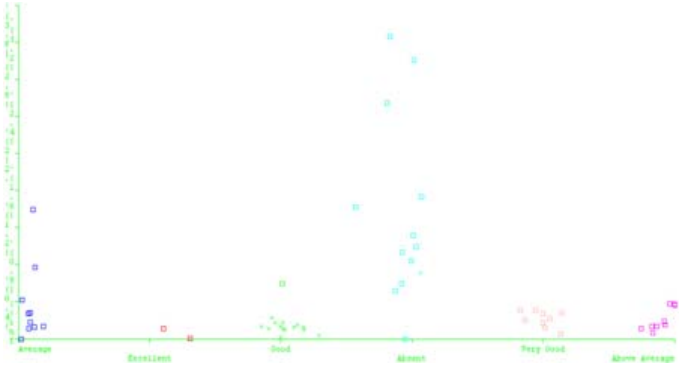


*f) Attendance Vs Performance*



*Current Standing Arrears Vs Performance*

*g)*



*h) Overall Attributes Vs Performance*

*Figure: Clustering of each individual attributes based on their overall performance*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure (a) to (h) depicts the clustered pattern of each attributes with their performance. The proposed method performs better clustering process on all the attributes and it can be easily observed from the cluster visualization.

*Table: Performance Evaluation Of The Proposed Fuzzy K-Mediods With K-Means And K-Mediods*

| | K-means | K-Mediods | Fuzzy K-Mediods |
|---|---|---|---|
| Correctly Clustered Instance | 40 | 47 | 52 |
| Incorrectly Clustered Instance | 18 | 11 | 6 |
| Accuracy | 69% | 81% | 90% |
| Precision | 75% | 87% | 94% |
| Recall | 72% | 86% | 92% |

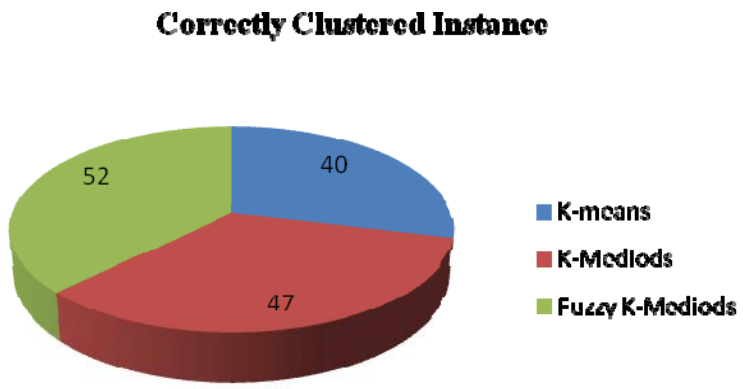



*Figure Performance Analysis Of K-Means, K-Mediods And Fuzzy K-Mediods Based On Correctly Clustered Instances*

The correctly classified instances of Fuzzy K-mediods in 52, K-means is 40 and K-Mediods is 47 out of 58 instances. The fuzzy K-medidos clustering is considerably high because of assigning each instances to various cluster based on their membership value.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

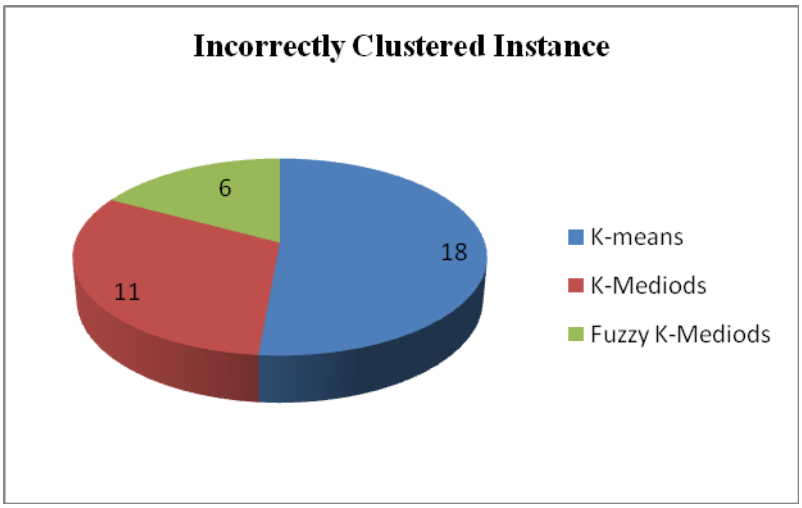



*Figure Performance Analysis Of K-Means, K-Mediods And Fuzzy K-Mediods Based On Incorrectly Clustered Instances*

The Incorrectly classified instance of Fuzzy K-Mediods is 6, K-means is 18 and K-Mediods is 411 out of 58 instances. The fuzzy K-Mediods clustering is considerably low because of assigning each instances to various cluster based on their membership value.

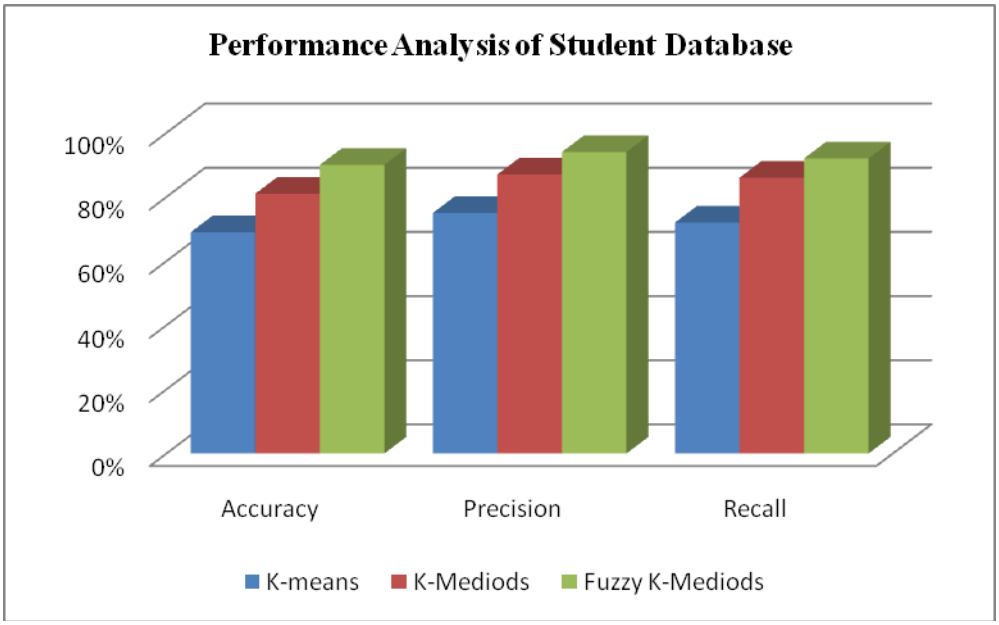



*Figure Performance Analysis Of K-Means, K-Mediods And Fuzzy K-Mediods Based On Accuracy, Precision And Recall*

The figure depicts the metrics of Accuracy, Precision and Recall of K-means, K-Mediods and Fuzzy K-Mediods. The True positive rate is high in Fuzzy K-Mediods and very low in K-means. The accuracy of clustering instances rely on the number of clusters assigned and knowledge in assigning the vague instances in appropriate clusters. The fuzzy K-Mediods overcomes it by assigning the membership value of a data instance to all the clusters created and the one which holds the highest membership cluster value is considered to be prominently belonging to it.