International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 8 Issue 01 Ver. II || Jan 2019 || PP 22-30

An Innovative K* Clustering Algorithm on Systematic Transformation of Asynchronous Regions for Estimating Education completion performance

S.N.Ali Ansari[1], Dr.Srinivasa Rao V[2], Dr.V.Srinivas[3]
[1](Research scholar, Rayalaseema university, Kurnool, India)
[2](Dept of CSE, V.R.Siddhartha Engineering college, Vijayawada, India)
[3](Dept of CSE, VSM College of Engineering, R.C.Puram, India)
Corresponding Author: S.N.Ali Ansari

Abstract: *In present days, the educational institutions maintain volumes of data of the students. The amount of data stored in educational databases is rapidly increasing because of the increase in awareness and application of data science in the field of higher and professional education system. we can mine the hidden knowledge in the available databases for generating various analytical reports for proper decision making. This Proposal is designed to present and justify the capabilities of data mining in data science environment. The main contribution of this proposal is the Estimating Education completion performance based on Systematic Transformation of Asynchronous Regions using K* Clustering Algorithm. The data stored in the Institution Education System (IES) from 2012 to 2016 will be used to perform an analysis of study on database for final result for this thesis. The confidentiality of data is also maintained. The final outcomes will be shown that most of the students belong to the cluster which needs motivation and remedial coaching for improving their educational capabilities. The dataset can be improved by including data of students currently enrolled during 2017-18 also. The result obtained can be used as a decision support component for any educational system. The WEKA/Python software will be used to Estimate Education completion performance in educational institutions using Systematic Transformation of Asynchronous Regions (STAR) K* clustering algorithm.*

Keywords: *Algorithm, clustering, Database, Data mining, hidden knowledge, K*, Python, WEKA*

Date of Submission: 30-12-2018 Date of acceptance: 15-01-2019

I. Introduction

Since ancient days, Education system is the core system for any economic growth of the society [1]. Similarly, a literally educator, with young mind and with his/her knowledge, is the main resource of improvement of the economic growth of our society. Hence, it is crucial element for education to be shaped with in accordance to the exact needs of the society and any organization. Today, higher/technical/professional education institutions such as universities are facing the problem of maintaining student retention rate which is related to education completion rates. Institutions with higher irregular student's retention rate tend to have higher graduation performance rate within four years or three years. Since were the most vulnerable to low student retention at all higher education institutions, early identification of irregular or vulnerable students who are prone to drop their courses is crucial for the success of any retention strategy. This would allow education institutions to undertake timely and proactive measures. Early identification of at-risk students can be the recipient of academic and administrative support to increase their chance of staying in the course and eventually complete the program. The ability to discover hidden information from institution databases particularly on enrolment data is very important in an educational institution. Being able to monitor the progress of student's academic performance is a critical issue to the academic community of higher learning. It is a long term goal of higher educational institutions to increase retention of their students. Institution completion is significant for students, academic and administrative staff. The importance of this issue for students is obvious: graduates are more likely to become entrepreneurs or get good jobs and earn more than those who dropped out from the institutions. With the help of data mining with data science, which is an essential process where intelligent machine learning methods are applied in order to extract data patterns, it is possible to discover the key characteristics from the students' records and possibly use those characteristics for future prediction of student retention rate. The K* clustering technique will be employed in order to discover data pattern with dynamic number of clusters and clusters can be formed in the form of Systematic Transformation of Asynchronous Regions. The students will be the first beneficiary of any improvement on the present policies. Faculty members and advisers will be properly informed of the status of their students. This study will provide the community about the factors affecting the institution completion giving them an idea on the value of the grades in high

school and the scores in the college admission test as two of the requirements for admission. Once admitted, performance in the freshman year is also a determining factor in their desire to have a college diploma to have a better job and eventually better lives. This study will also help the University in providing good educational services from the time they enrolled for the new course until their last semester of reside in the institution to complete their degree. The only available data about the students in the institutions' database is the information they supplied in their enrolment form. It is a challenge on the part of the institution administrators and academic planners to update records of students with relevant information that will aid in any academic related decision that may be needed in the future.

1.1 Research Objectives: The main objective of this study is to explore the enrolment data that may impact the study outcome of students. Specifically, the enrolment data were used to achieve the following objectives:

1. To build Education completion performance model based on K* clustering data mining technique on the basis of identified attributes.
2. To discover the overall distribution pattern and correlation among data attributes; and
3. To determine the significant attribute that contributed to the Education completion performance model.4. The performance of algorithm can be increased, if the number of clusters is reduced dynamically.

II. Background And Related Work

Conducted a study and investigated academic performance among college/university students living in three different residential environments dormitory, apartment house and commuter [2]. Analysis is conducted on sample of 471 students attending a Four-year educational institution. Significant differences are found between commuters and apartment residents. Controlling several background characteristics, being a commuter student positively influences Grade Point Average (GPA) in comparison to living student in apartment adjacent to campus. Age, educational objectives, and race also have significant effects on academic performance in higher educational institutions.

Several factors have been identified as hampering academic work and pupils' performance in public schools and colleges [3]. For instance, Etsey et al. (2004) in their study of some private and public schools in Ghana revealed that academic performance is better in private schools due to more effective supervision of work. Thus, effective supervision improves the quality of teaching and learning in the classroom (Neagley and Evans, 1970). Also, the attitude of some public school teachers and authorities to their duties does not engender good learning process for the pupils. Some teachers leave the classroom at will without attending to their pupils because there is insufficient supervision by circuit supervisors. This lack of supervision gives the teachers ample room to do as they please. Another factor is lack of motivation and professional commitment to work by teachers. (Young, 1989). This produces poor attendance and unprofessional attitudes towards pupils by the teachers, which in turn affect the performance of the pupils academically (Lockheed and Verspoor, 1991). Apart from all the aforementioned, most public schools lack adequate infrastructure and educational facilities. For instance, reading and learning materials are mostly hardly available, especially in rural areas. Also the size of each class forms a critical determinant of pupils'academic improvement and performance (Cochran-Smith, 2006). For example, Kraft (1994) in his study of the ideal class size found that class sizes above 40 pupils have negative effects on pupils'academic achievement. This is because of the possible differences in interests and abilities of pupils, particularly in commanding attention in class (Asiedu-Akrofi, 1978).

The study conducted by [4] examined the validity of High-school grades in predicting student success beyond the freshman year. The results showed that high-school grade point average (HSGPA) is consistently the best predictor not only of freshman grades in college, but of four-year college outcome as well. The study tracked four-year college outcomes, including cumulative college grades and graduation, for the same sample in order to examine the relative contribution of high school record and standardized tests in predicting longer term college performance.

Key findings showed that HSGPA is consistently the strongest predictor of four-year college outcomes for all academic disciplines. The predictive weight associated with HSGPA increases after the freshmen year, accounting for a greater proportion of variance in cumulative fourth-year than first-year college grades. Other factors such as standardized test, school academic performance index, socio-economic status and parent's education were considered by only to concede to HSGPA as a valid factor for predicting success beyond freshman year. A model was developed using a structural equation modeling to explain education performance of undergraduate students [5]. The independent variables were perceived institutional support, academic self efficacy, institutional commitment, classroom learning environment and social support. The conclusion reached from the analysis is that the learning environment is a moderately powerful but indirect influence on student college completion intention. Social support along with perceived institutional support contributes to a student's intention to complete education. Academic self-efficacy also plays a smaller yet significant role in student's education completion intention.

conducted a study on the student performance by means of rule-based classification method on 17 attributes, it was found that the factors like students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and students family status were highly correlated with the student academic performance[6].

The application of Data Mining in the education sector was explored by [7]. The study takes the performance of students in their examination and their presence in the classroom and finds a relation in them. The observed relation helps in identifying the group of students where the extra attentions are required. The study was carried out using K-means method of cluster analysis. A study conducted by [8] revealed that preadmission scholastic assessment test (SAT) scores and high school record are significant predictors of graduation. The correlations observed were moderate and lower than the correlations of admission credentials with cumulative GPA. Other predictors and criteria of success which are non-academic and which clearly influence persistence in college are financial status, health and student personality. A case study was presented on educational data mining to identify up to what extent the enrolment data can be used to predict student's success [9]. The algorithms Chi-squared Automatic Interaction Detector (CHAID) and Classification and Regression Tree (CART) were applied on student enrolment data of information system students of open polytechnic of New Zealand to get two decision trees classifying successful and unsuccessful students. The accuracy obtained with CHAID and CART was 59.4 and 60.5 respectively. An intelligent student advisory framework in the educational domain was developed by [10]. They classified the students into the suitable department using C4.5 algorithm. They also clustered the students into groups as per the suitable education tracks using k-means algorithm. They combined the results that came out from classification and clustering operations to predict more results. A case study was presented to prove the efficiency of the proposed framework. Students data collected from Cairo Higher Institute for Engineering, Computer Science and Management during the period from 2000 to 2012 were used and the results proved the effectiveness of the proposed intelligent framework. Data mining approach to differentiate the predictors of retention among the freshmen enrolled at Arizona State University [11]. Using the classification tree based on an entropy tree-splitting criterion they concluded that cumulated earned hours was the most important factor contributing to retention. Gender and ethnic origin were not identified as significant. Conducted a study to analyze students' results based on cluster analysis and used standard statistical algorithms to arrange their scores according to their level of performance [12]. K-means clustering algorithm was implemented. The model created was an improvement of the limitation of existing methods developed by Omelehin using fuzzy logic. In their study presented a hybrid procedure based on decision tree of data mining method and data clustering which will enable the academicians to predict student's GPA based [13].

III. Work Done/Contribution

3.1 Framework of the Study: The framework of the study was based on the Knowledge Discovery Process (KDP) on databases illustrated by [14]. The KDP figure was modified to suit the objectives of the study. The modified version was presented on Fig. 1 following the steps from preprocessing wherein noisy and irrelevant data were removed, selection and transformation where data relevant to the analysis task were retrieved from the database and further transformed or consolidated into forms appropriate for mining, data mining where K* clustering were applied in order to extract data patterns in a systematic manner, interpretation and evaluation where the truly interesting patterns representing knowledge based were identified and knowledge presentation where visualization and knowledge presentation techniques were used to present the mined knowledge to the management of the institutions.

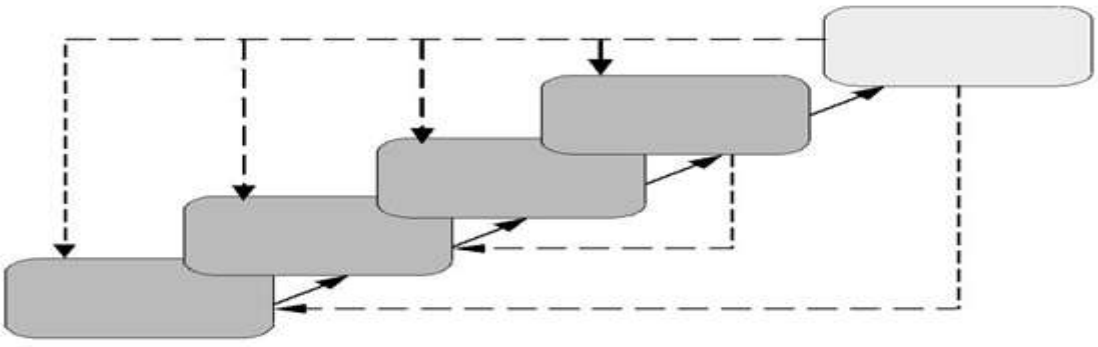

Fig. 1. The steps of extracting knowledge from data.

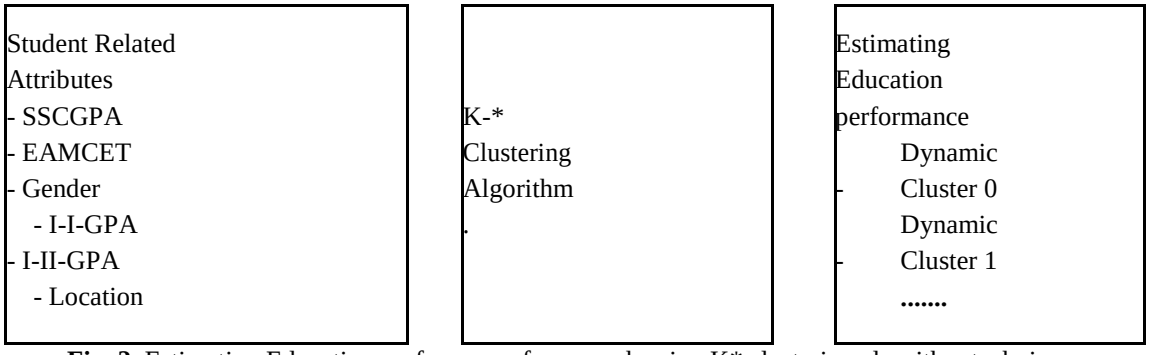

Fig. 2. Estimating Education performance framework using K* clustering algorithm technique.

3.2 Estimating Education performance Framework:

Data mining is just a part of the whole framework of the study. Fig. 2 shows the estimating education performance process framework as its major elements used in this study. The information stored in the Institution Education System were analyzed to be able to extract suitable dataset for the study. The dataset was produced after data pre-processing. This served as input to the data mining tool for the application of the selected K* clustering algorithms to estimate the education performance in the institution. Two clusters were produced after the process.

The education performance was evaluated based from their accuracy consistent with the results obtained from training dataset. The number of required clusters can be generated by using input dataset. Cluster number represents groups of student related or similar with each other. The knowledge discovered can then be used for decision making.

3.3. Methodology/Data Mining Process:

In this study normal K-means Algorithm,, the Weka data mining tool and Python programming language with Pandas are used, which offers different data mining techniques for various kinds of datasets. The Weka Knowledge Explorer (WKE) is an easy to use Graphical User Interface (GUI) that binds the power of the Weka software. The major Weka packages are Filters, Classifiers, Clusters, Associations, Attribute Selection and Visualization tool, which allows datasets and the predictions of Classifiers and Clusters to be visualized in two or three dimensions. The workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling together with graphical user interfaces for easy access to this functionality. Weka was primarily designed as a tool for analyzing data from agricultural domains. It is now used in many different application areas, in particular for educational purposes and research [15].

K* clustering algorithm was selected to analyze the dataset extracted from the Institution Education System database. Fig. 3 shows the process of how STAR (Systematic Transformation of Asynchronous Regions) K* clustering work. A clustering algorithm attempts to find natural groups of components (or data) based on some similarity in systematic way of identification boundary of each region. Also, the clustering algorithm finds the centroid of a group of data sets. To verify cluster membership in systematic way, most algorithms estimate the distance between a point and the cluster centroids. The output from a clustering algorithm is basically a statistical narration of the cluster centroids with the number of components in each cluster.

3.4 Proposed Method: The K-means algorithm finds the predefined number of clusters. In the practical scenario, it is very much essential to find the number of clusters for unknown dataset on the runtime and data points should be aligned into corresponding clusters. The fixing of number of clusters may lead to poor quality clustering. The proposed method finds the number of clusters on the run based on the cluster quality output. In the first part of algorithm, the number of clusters is found. Then assign data points to initial cluster. Then, find an effective and useful initial centroid. In these two parts; there have some loops and its calculation but these make a better clustering rather than other clustering methods. For make algorithm more intelligent, some procedure must add. Intelligence of algorithm will help to determine the number of cluster and which points are the initial centroids. In last portion of modified algorithm, feasible data points which have chance to change current cluster and move to new cluster. A short list of points for calculation is prepared. It minimizes calculation, that's why it was capable to save time on behalf of original standard K-means algorithm.

In step 1, Calculate the required number of clusters based on dataset and a set of predicates dynamically and Apply standard K-means algorithm, or to calculate the required number of clusters by creating clusters dynamically, Apply the below steps. In step 2, calculate the equation and get a concept about the number of clusters. From step 3, assign a value in Clno for maintaining cluster number. By step 4, 5 and 6,

algorithm assigns a new cluster and assigns data points to this cluster. From step 7, algorithm finalizes the initial cluster's member data points, and gets decision to start a new cluster. Step 8; find the initial centroids for clusters. By help of step 8, step 9 finalized a stable cluster and centroid. Step 9 finds The closest centroid for each data points in a systematic way of identifying boundary curve and allocate each data points Cluster. Calculate new centroid for each R_m. Every time, the Asynchronous regions and corresponding data points are aligned into respective clusters. In step 10, 11, 12 and 13, there have tricks to find feasible data points those have chance to change current cluster. Calculate interval's points. Normally other points don't move clusters. For this step, algorithm saves a lot of time. It minimizes a lot of calculation. To get decision, step 13 is used. In this step; algorithm take decision about the algorithm continue or all clustering is finished.
At initial stage, data points of clusters can be scattered, during processing of this algorithm, the data points of asynchronous regions can be transformed in a systematic manner. By comparing boundaries of each region, the data points can be aligned to corresponding clusters.
The K* clustering algorithm is as follows:

Input:
Input: I = n_1, n2, n3,---n // Set of n number of data points
Output: A set of R (Systematic Regions) Clusters. // Number of desired Clusters.

Method:
Step 1. Calculate the required number of clusters based on dataset and a set of predicates dynamically and Apply standard K-means algorithm, or to calculate the required number of clusters by creating clusters dynamically, Apply the below steps.
Step 2. Calculate R = sqrt (n/2) .
Step 3. Allocate Clno = 1 where Clno indicates cluster number.
Step 4. Find the closest pair of data point from I. Move those points to new set N_{Clno}.
Step 5. Find the closest point of N_{Clno} and move it to N_{Clno} from input I.
Step 6. Repeat step 5 until the number of elements of N_{Clno} reach (n/R)*Clno.
Step 7. When, N_{Clno} are full and the number of elements of Input I! = 0. Increment the value of Clno.
New Clno = Clno + 1. Repeat step 4.
Step 8. The center of gravity of each N_{Clno}. Those are the initial centroids C_j and elements of N_{Clno} are the elements of R_m. Step 9.k= k+1 go to Step 1.
Step 9. Find the closest centroid for each data Points in a systematic way of identifying boundary curve and allocate each data points Cluster. Calculate new centroid for each R_m. Every time, the Asynchronous regions and corresponding data points are aligned into respective clusters.
Step 10.Calculate the largest distance D_L (largest distanced data point from each centroid of each cluster).
Step 11.Get the data points in the interval of (D_L * 4/9) to D_L .
Step 12.Find the closest centroid for those points and allocate them to closest centroid cluster.
Step 13.Find new centroid for each R_m.If, any change in any R_m.Repeat step 10. Otherwise go to end of the algorithm Step 14.
Step 14.Stop

Field of Modification: K-means algorithms main limitation fields are; fix the value of K, Make an initial centroid. Lot of researcher worked these fields particularly. Our point of view is modified k-means as; it can overcome all limitations. Our concern of modification is combined all limitation and solved them for maximum output. We solve the number of cluster problem. Algorithm can determine own its number of cluster by working data points. Our expression of determine number of cluster is $R \cong \text{sqrt}(n/2)$. Here, R= number of clusters. n= number of data inputted. It's the initial concept of Number of cluster. We fix the number of clusters in next equation. Compute the closest pair and keep them to N_{Clno}, and delete them from input set I. Then, fill N_{Clno} by calculate the nearest of those pair when any data goes to any N_{Clno} then it was deleted from I.

When N_{Clno} is full we increase the value of Clno and continue same process as loop until in input I have any data point left. Closest pair points mean value is initial centroids. Assign closest points of initial points into Rm. Those are initial clusters. Calculate the center of gravity to find the centroids. When we get each Rm center of gravity, those are the centroids of each Rm. Now we have initial clusters and centroids.

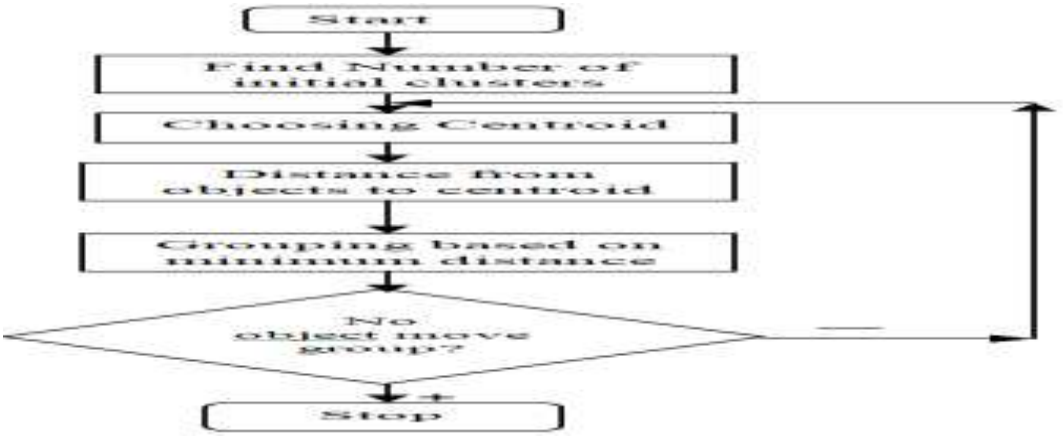

Fig. 3. K-means clustering process flowchart

A total of 164 academic records of newly admitted students enrolled in various programs in the university in the institution in the year 2016-2017 were taken as a sample from a total population of 1053 newly admitted students. Students from other institutions who transferred to the institution were not included in the study. Description of variables and their data types were presented in Table I. Fig. 4 shows enrolled and passed students chart from 2012-13 to 2015-16.

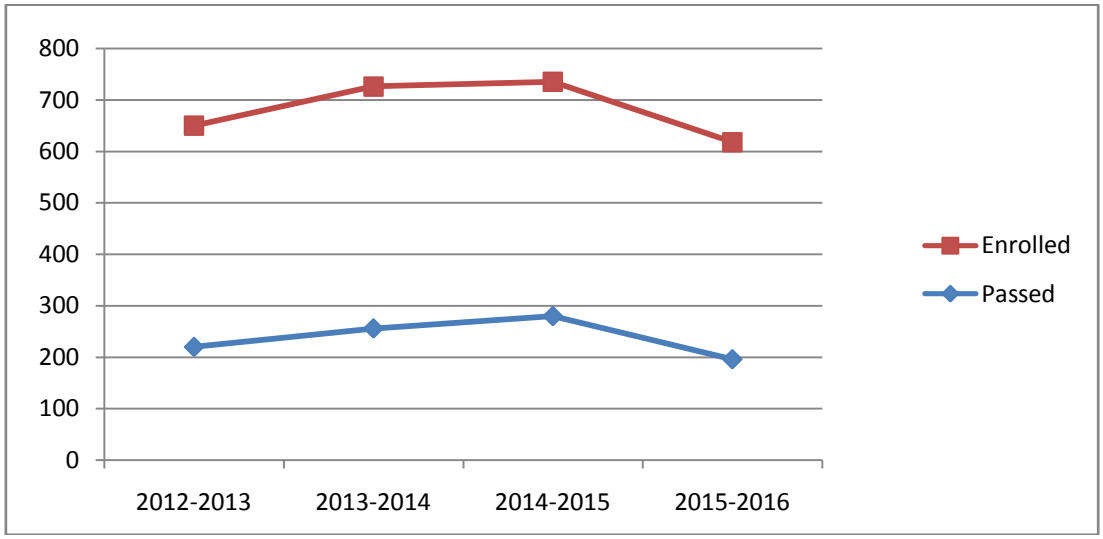

Fig. 4 Enrolled and completed indicator

The domain values of the attributes used were defined as follows:

SSCGPA-Secondary School Certificate Grade Point Average. It is the general weighted average of the student in in high school.

EAMCET- Common Entrance Test. It is a standardized test given to student who intends to admit in the Engineering institution.

GENDER- Student's gender. It is the category of student whether male or female. M represents male while F represents Female.

I-I-GPA-Grade Point Average in first year first semester. This is the average grade of student while in first year first semester. The value range from 1.0 to 10.0 where 10.0 is the highest grade a student can get and 1.0 as the lowest grade.

I-II-GPA-Grade Point Average in first year second semester. This is the average grade of student while in first year second semester. The value range from 1.0 to 10.0 where 10.0 is the highest grade a student can get and 1.0 as the lowest grade.

Table I: student related attributes and data types

SNo	Variable	Description	Data Type
1	SSCGPA	Secondary School Certificate GPA	Numeric
2	EAMCET	Engineering Entrance Test	Numeric
3	GENDER	Student's Gender M or F	Nominal
4	I-IGPA	I-I Semester GPA	Numeric
5	I-II GPA	I-II Semester GPA	Numeric
6	Location	Area may be Rural or Urban	Nominal

IV. Results And Discussions

The objective of the study is to estimate student performance in the institutions using K* clustering technique on the basis of identified attributes which were SSCGPA, EAMCET, GENDER, I-I-GPA, and I-II-GPA.
The result is shown below:

Fig. 5 shows the output of K* clustering algorithm when executed. Two clusters were formed after 3 iterations. Cluster 0 contains 74 instances or 24%, Cluster 1 contains 43 instances or 14%, Cluster 2 contains 115 instances or 38%, Cluster 3 contains 72 instances or 24%. This analysis showed that in Cluster 2 and Cluster 3 contains high GPA in both I-I and I-II semesters out of 304 students.

```
Start time--- 1542073050.878227 seconds ---
Converged after 3 iterations
End time--- 1542073050.9250941 seconds ---
Time taken--- 0.04686713218688965 seconds ---
Cluster:  0  points= 74  Percent is= 24 2
Cluster:  1  points= 43  Percent is= 14 2
Cluster:  2  points= 115  Percent is= 38 2
Cluster:  3  points= 72  Percent is= 24 2
```

Fig. 5. Output of K* clustering algorithm

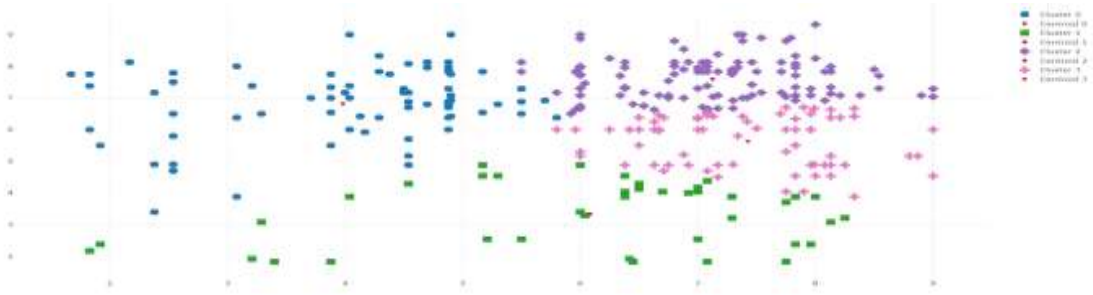

Fig. 6. Graphical representation of clustered instances

Fig. 6 shows a graphical representation of clustered instances. Cluster 0 is in the lower region of the axes which is compose of 74 students while Cluster 3 is on the upper region of the axes which is compose of 72 students.

The results showed that students from Cluster 3 and Cluster 2 are more likely to complete college on time than those students in Cluster 0 and Cluster 1. Students in Cluster 0, Cluster 1 are considered at-risk students. They are the students who are more likely to drop or stay longer in the university to finish education. The Python language was utilized to discover the overall distribution pattern and correlation among data attributes. Fig. 7 shows the correlation matrix.

It shows that the EAMCET Entrance Test got the highest correlation coefficient among the 4 variables which implied that 1-2 GPA is strongly correlated and highly significant.

Description of Sample data

	SSCGPA	EAMCET	1-1-GPA	1-2-GPA
count	304.000000	304.000000	304.000000	304.000000
mean	8.429770	84924.476974	6.283520	6.337664
std	1.079632	78605.321526	1.711433	1.665783
min	0.950000	431.000000	1.670000	1.830000
25%	8.000000	49397.500000	5.170000	5.170000
50%	8.700000	78814.000000	6.750000	6.700000
75%	9.200000	114438.000000	7.540000	7.700000
max	10.000000	989879.000000	9.000000	9.320000

Correlation of Sample data

	SSCGPA	EAMCET	1-1-GPA	1-2-GPA
SSCGPA	1.000000	-0.040208	-0.081212	0.064914
EAMCET	-0.040208	1.000000	0.026663	0.066478
1-1-GPA	-0.081212	0.026663	1.000000	0.042035
1-2-GPA	0.064914	0.066478	0.042035	1.000000

>>>

Fig. 7. Description & Correlation matrix.

V. Conclusion

The study examined the available enrolment data of students in the university's database. Based on result from K* clustering, 38% of 304 freshmen enrolled were considered at-risk of not completing their graduation on time while 62% has a greater chance of completing graduation. For estimating education completion performance, it can be concluded that score in the 1-2 GPA is a significant factor in determining education completion performance as it gets a correlation coefficient of +0.066. This paper is an endeavor in providing the new method of taking advantage of the available data for the improvement of educational process via data mining technology. The main idea is to come up with a estimating education completion performance which can be used to improve the decision making processes for remedial coaching and managing and monitoring attendance of students.

VI. Future Work

Future researchers may use the model to identify the existing area of research in the field of data mining in higher education. They may use additional predictor variables related to students and institution that may have effect on the retention and education completion performance of students in the institutions like colleges, universities. The inclusion of the records of currently enrolled students is highly recommended to monitor their progression and for early intervention for those who may be considered at-risk. The development of decision support system may also be undertaken for a more efficient monitoring and effective decision making

References

[1]. Adekemi J. Oluwadare. "Towards a Knowledge-Based Economy: Challenges and Opportunities for Nigeria," International Conference on African Development Issues (CU-ICADI) 2015: Social and Economic Models for Development Track.

[2]. Michael Delucchi, "Academic performance in college town", Education Vol.114 No,1 p96-100.

[3]. Okyerefo ,Daniel Yaw Fiaveh and Steffi Naa L. Lamptey, "Factors prompting pupils' academic performance in privately owned Junior High Schools in Accra, Ghana" International Journal of Sociology and AnthropologyVol. 3(8) ,pp. 280-289, August 2011.

[4]. Laura Pagani and Chiara Seghieri, "Predictive Validity of High School Grade and other Characteristics on Students' University Careers using ROC Analysis," Developments in Applied Statistics Anuˇska Ferligoj and Andrej Mrvar (Editors) Metodoloˇski zvezki, 19, Ljubljana: FDV, pp.197-204,2003.

[5]. Darrin Thomas, " Factors that Influence College Completion Intention of Undergraduate Students," TheAsia-Pacific Education Researcher June 2014.

[6]. M.S. Mythili, Dr. A.R.Mohamed Shanavas, "An Analysis of students' performance using classification algorithms," OSR Journal of Computer Engineering (IOSR-JCE)e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 16, Issue 1, Ver. III (Jan. 2014), PP 63-69.

[7]. S. P. Singh, B. K. Sharma, and N. K. Sharma, "Use of clustering to improve the standard of education system," International Journal of Applied Information Systems, vol. 1, no. 5, pp. 16-20. February 2012.

[8]. N. W. Burton and L. Ramist, "Predicting success in college: SAT studies of classes graduating since 1980," College Entrance Examination Board, New York, 2001.

[9]. Jai Ruby, Dr. K. David, "Predicting the Performance of Students in Higher Education Using Data Mining Classification Algorithms -A Case Study," International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 2 Issue XI, November 2014.

[10]. Heba Mohammed Nagy, Walid Mohamed Aly, Osama Fathy Hegazy, "An education data mining system for advising higher education students," International Journal of Computer, Information Science and Engineering, vol. 7, no. 10, 2013.

[11]. Chong Ho Yu, Samuel DiGangi, Angel Jannasch-Pennell, Wenjuo Lo, Charles Kaprolet , "A data-mining approach to differentiate predictors of retention," presented at the EDUCAUSE Southwest Conference, Austin, Texas, USA 2007.

[12]. O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of K-means clustering algorithm for prediction of students" academic performance," International Journal of Computer Science and Information Security, vol. 7, no. 1, 2010.

[13]. M. H. I. Shovon and M. Haque, "An approach of improving students academic performance by using K-means clustering algorithm and decision tree," International Journal of Advanced Computer Science and Applications, vol.3, no. 8, 2012.

[14]. J. Han and M. Kamber, "Data mining: concepts and techniques," Simon Fraser University, Morgan Kaufmann publishers.
[15]. S. K. Yadav, B. Bharadwaj, and S. Pal, "Data mining applications: a comparative study of predicting students performance," International Journal of Innovative Technology and Creative Engineering, vol. 1, no. 12.
[16]. Shailendra Singh Raghuwanshi, PremNarayan Arya,"Comparison of K-means and Modified K-mean algorithms for Large Data-set", International Journal of Computing,Communications and Networking, Volume 1, No.3, November –December 2012.

S.N.Ali Ansari" An Innovative K* Clustering Algorithm on Systematic Transformation of Asynchronous Regions for Estimating Education completion performance" International Journal of Engineering Science Invention (IJESI), vol. 08, no. 01, 2019, pp 22-30