Academic Decision Support System for College Completion Model

Allen M. Paz, Bobby D. Gerardo, Bartolome T. Tanguilig III

Abstract— **Universities need to have extensive capabilities in order to analyze students' achievement levels which will help in making appropriate academic decisions. Conversely, academic decisions will result in changes in academic performance which need to be assessed periodically and over spans of time. In this work, the college completion model based on k-means clustering algorithm was utilized in the development of the proposed academic decision support system (DSS). The system utilized data from the university database while the client front-end ensures adequate presentation so as to reveal significant details and dependencies. The system can be used to automate the decision making process of administrators aiming to decrease the high rate of academic failure among students. A real case study in Isabela State University is presented using a dataset collected from 2009-2013.**

Keywords—decision support system, k-means clustering, college completion, data mining

I. Introduction

The ability to discover hidden information from university 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. College completion is significant for students, academic and administrative staff. It means earning a college degree on time or staying longer than the expected time to complete. The importance of this issue for students is obvious: graduates are more likely to find decent jobs and earn more than those who dropped out.

With the help of data mining which is an essential process where intelligent 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

Allen M. Paz
Isabela State University
Philippines

Bobby D. Gerardo, Ph.D.
West Visayas State University
Philippines

Bartolome T. Tanguilig III, Ph.D.
Technological Institute of the Philippines
Philippines

future prediction. K-means clustering technique was employed in order to discover pattern.

The research presented here was inspired by the evolving reforms in the higher education system in the Philippines aimed at improving the performance of public universities in changing economic condition. Universities and colleges are required to submit reports on the number of graduates per school year. Statuses of those who either dropped or stay longer than expected time to graduate were not known. Manual method of reviewing student records was observed. The problem with this method is that it takes time to produce accurate report, the factors affecting small number of graduates are not known and that it is too late to identify those who dropped and staying longer to the program they enrolled in. Thus, the researchers came up with an academic decision support system for college completion. Decision support system (DSS) is a knowledge-based information system to capture, handle and analyze information which affects or is intended to affect decision making performed by people in the scope of a professional task appointed by a user [2].

The development of academic decision support system for college completion will benefit the students and the academic planners. Students will be properly informed about their academic standing in the early part of their studies and academic planners will be able to design better strategies for early intervention. This study will also help the school in providing better educational services from the time they registered for the first time until their last semester of stay in the university to complete their degree. The ability to monitor students' progression is important for an academic institution to be able to help in the nation builds its human capital to enable it to actively participate in the global economy [1].

The goal of our research is to contribute to the next generation DSS with incorporated data mining and knowledge discovery functionalities. Major tasks of the system were defined as follows:

1. Relevant input data has to be extracted from university's database to provide data needed for querying and clustering
2. A sound college completion model has to be designed.
3. The proposed model has to be implemented in a user-friendly DSS interface.

The user friendliness, usability, expandability and cost effectiveness of the system were taken into consideration.

The present paper is an extension of our latest study and publication [9]. The development of college completion model based on k-means clustering was presented using the identified attributes extracted from the university's database. The overall distribution pattern among data attribute were

presented and the significant attribute that contributed to the college completion was identified.

Our contribution is basically twofold: 1) development of college completion model based on k-means clustering, and 2) to design an academic decision support system. The paper is structured as follows: Section 2 presents the review of related studies followed by the methodology in Section 3; in Section 4 a discussion of the results of implementation was presented. We conclude by a summary of our contribution and proposals for future work.

II.   Related Studies

Many researchers have contributed to the field of educational data mining and decision support system in higher education.   In this section, the researchers will give an overview on a few representative works.

Chau and Phung [3] in their study stated that education always plays an important role in building up every country around the world. According to them, educational decision making support is significant to students, educators, and educational organizations and the support will be more valuable if a lot of relevant data and knowledge mined from data are available for educational managers in their decision making process. They proposed a knowledge-driven educational decision support system. The knowledge-driven decision support is helpful for educational managers to make more appropriate and reasonable decisions about student's study and further give support to students for their graduation. A waste of effort, time, and money can be avoided accordingly for both students and educators through the proposed system.

Deniz and Ersan [4] presented different ways in which student performance data can be analyzed and presented for academic decision making. They demonstrated the usefulness of an academic decision-support system (ADSS) in evaluating huge amounts of student-course related data. In addition, they presented the basic concepts used in the analysis and design of a specific DSS software package which is called the Performance based Academic Decision Support System (PADSS).

The study conducted by Feghali, Zbib and Hallal [5] attempts to solve a technology-based "last mile" problem by developing and evaluating a web-based decision support tool – the Online Advisor, that helps advisors and students make better use of an already present university information. Their study showed that 79% of users stated that they were satisfied with the Online Advisor, 90% rated the Online Advisor as effective and efficient and more than 75% rated the Online Advisor as useful and helpful.

According to Kotsiantis [6] the use of machine learning techniques for educational purposes or educational data mining is an emerging field aimed at developing methods of exploring data from computational educational settings and discovering meaningful patterns. The stored data can be useful for machine learning algorithms. Students' key demographic characteristics and their marks in a small number of written assignments can constitute the training set for a regression method in order to predict the student's performance. A prototype version of software support tool for tutors has been constructed.

Nagy, Aly and Hegazy [7] proposed a "Student Advisory Framework" that utilizes classification and clustering to build an intelligent system. The system can be used to provide pieces of consultations to a first year university student to pursue certain education track where he/she will likely to succeed in. According to them, one of the main reasons for high failure rate is the incorrect selection of the student's department/section. The framework acquires information from the datasets which stores the academic achievements of students before enrolling in a certain department. After acquiring all the relevant information, a new student can challenge the intelligent system to receive a recommendation of a certain department in which he/she would likely succeed. Results have proved the efficiency of the suggested framework.

Vinnik and Scholl [10] proposed a methodology for assessing educational capacity and planning its distribution and utilization in universities. They integrated educational data mining and knowledge discovery in their proposed method. The DSS supports the administrative task of planning the university's educational capacity in terms of the number of students its courses can accommodate under the specified constraints. Decision-makers were able to evaluate various strategies and generate forecasts by means of simulating with the input data. According to them, when the policy makers applied the system it has resulted in significant acceleration in planning procedures, raised the overall awareness with respect to the underlying methodology and ultimately enabled more efficient academic administration.
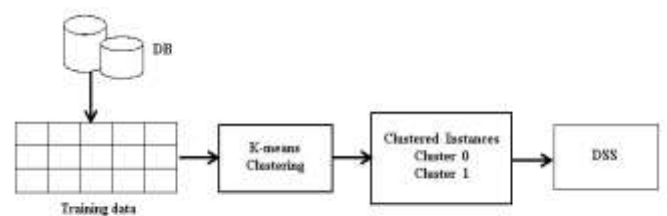
III.   Methodology

Figure 1. Framework of the college completion model for decision support system

Figure 1 shows the proposed design for the clustering system to build college completion model. The information stored in the Student Accounting Registration and Information System were analysed to be able to extract appropriate 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-means clustering algorithms (Figure 2). Two clusters were produced after the process.

International Journal of Advances in Computer Science & Its Applications – IJCSIA
Volume 4: Issue 4   [ISSN 2250-3765]

Publication Date : 27 December,2014

K-means Clustering Algorithm

Suppose that a dataset of n data points x_1, x_2, ..., x_n such that each data point is in R^d, the problem of finding the minimum variance clustering of the dataset into k clusters is that of finding k points $\{m_j\}$ (j=1, 2, ..., k) in R^d such that

$$\frac{1}{n} \sum_{i=1}^{n} [\min d^2, (x_i, m_j)] \tag{1}$$

is minimized, where $d(x_i, m_j)$ denotes the Euclidean distance between x_i and m_j. The points $\{m_j\}$ (j=1, 2, ...,k) are known as cluster centroids. The problem in (1) is to find k cluster centroids, such that the average squared Euclidean distance (mean squared error, MSE) between a data point and its nearest cluster centroid is minimized.

The k-means algorithm provides an easy method to implement approximate solution to (1). The reasons for the popularity of k-means are ease and simplicity of implementation, scalability, speed of convergence and adaptability to sparse data [8].

The k-means algorithm begins at starting cluster centroids, and iteratively updates these centroids to decrease the objective function in (1). The k-means always converge to a local minimum. The particular local minimum found depends on the starting cluster centroids. The k-means algorithm updates cluster centroids until local minimum is found. k-means algorithm is shown below.

```
1   MSE=largenumber;
2   Select initial cluster centroids {m_j}_j^k =1;
3   Do
4       OldMSE=MSE;
5       MSE1=0;
6       For j=1 to k
7           m_j=0; n_j=0;
8       endfor
9       For i=1 to n
10          For j=1 to k
11              Compute squared Euclidean
                    distance d^2(x_i, m_j);
12          endfor
13          Find the closest centroid m_j to x_i;
14          m_j=m_j+x_i; n_j=n_j+1;
15          MSE1=MSE1+d^2(x_i, m_j);
16      endfor
17      For j=1 to k
18          n_j=max(n_j, 1); m_j=m_j/n_j;
19      endfor
20      MSE=MSE1;
        while (MSE<OldMSE)
```

Figure 2. K-means clustering algorithm.

A total of 174 academic records of freshmen students enrolled in various programs in the university in the school year 2009-2010 were taken as sample from a total population of 1053 freshmen students. Students from other institutions who transferred to the university with earned units were not included in the study. Descriptions of attributes and their data types were presented in Table 1.

TABLE 1. STUDENT RELATED ATTRIBUTES AND DATA TYPES

| Variable	Description	Data Type
HSGPA	High School Grade Point Average. This is the general average obtained in their last year in high school	Numeric
CET	College Entrance Test. This is the score obtained in the examination given by the university before entering college	Numeric
GENDER	Student's gender. This is the category of students whether male or female	Nominal
SCHOLAR	This is the category of student whether the student is enjoying scholarship or not	Nominal
FYGPA1	Grade point average in first year first semester	Numeric
FYGPA2	Grade point average in first year second semester	Numeric
PUBLIC_PRIVATE	This is the category of school last attended whether public or private	Nominal
ENROLLED_AFTER_HS	This is the category whether the student enrolled college right after finishing high school	Nominal
PREFERRED_COURSE	This is the category whether the student enrolled his/her priority course	Nominal
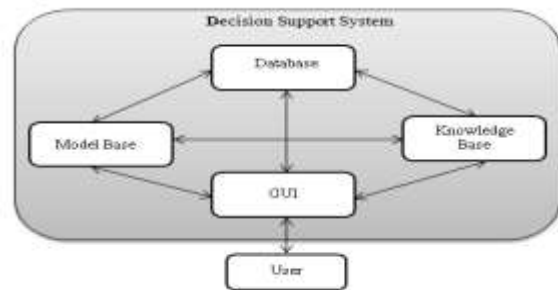
Figure 3. Components of decision support system

A DSS application contains five components: database, model base, knowledge base, graphical user interface (GUI) and User (Figure 3). The database stores the data, model and knowledge bases store the collections of models and knowledge, and the GUI allows the user to interact with the database, model base, and knowledge base.

The database and knowledge base can be found in the basic information system. The knowledge base contains simple search results for analyzing the data in the database. For example, the knowledge base may contain how many students are enjoying scholarship grants in a particular semester. A

decision support system is an intelligent information system because of the addition of the model base. College completion model based on k-means clustering algorithm is the content of the model in the proposed decision support system. This model allows the decision support system to not only supply information to the user but aid the user in making decision.

IV. **Results and Discussions**

Preprocessing

The major challenge was the preprocessing phase in which the entire input data has to be identified, collected, and integrated into a database. Since the previous research suggested to consider other variables available, the researchers considered whether the student enrolled college right after high school, is the course he or she enrolled in is his or her first priority course to take, and the type of previous school attended whether public or private [9].

The relevance or significance of the attributes were tested using chi-squared attribute evaluation found in Weka. The result using 10-fold cross validation evaluation mode is presented in Table 2. Based on the result of evaluation, it can be inferred that PREFERRED_COURSE got the highest rank among the identified attributes.

TABLE 2. ATTRIBUTE EVALUATION

Attribute	Average	Merit
PREFERRED_COURSE	156.6	+-0.49
FYGPA2	20.267	+-6.877
GENDER	5.585	+-1.842
SCHOLAR	0.635	+-0.506
ENROLLED_AFTER_HS	0.082	+-0.115
FYGPA1	0	0
HSGPA	0	0
PUBLIC_PRIVATE	0	0
CET	0	0

Modeling

The idea to enhance k-means algorithm was considered since in each iteration, the *k*-means algorithm computes the distances between data point and all centers and this is computationally very expensive especially for large datasets. For each data point, keep the distance to the nearest cluster. At the next iteration, compute the distance to the previous nearest cluster. If the new distance is less than or equal to the previous distance, the point stays in its cluster, and there is no need to compute its distances to the other cluster centers. This saves the time required to compute distances to $k-1$ cluster centers.

After using k-means clustering algorithm, Table 3 shows the output of processing: There were 174 instances evaluated and 9 attributes used. The resulting clusters are considered the model and will be the input for the decision support system. The initial starting points (random):

Cluster 0: 82.55.1.Y,N,41,F,Y,2.66,2.3
Cluster 1: 91.1,1,Y,Y,52,F,Y,1.87,1.92

The final cluster centroids are shown in Table 3. Cluster 0 contains 46 (26%) instances while cluster 1 contains 128 (74%) instances. Cluster 0 and 1 represent the college completion model. It is interesting to note that students who belong to cluster 0 enrolled course not in their priority. It implies that enrolling in the preferred program is a good motivation to complete college

TABLE 3. THE FINAL CLUSTER CENTROIDS AFTER K-MEANS CLUSTERING PROCESS

Attribute	Full Data (174)	Cluster #	
		Cluster 0 (46)	Cluster 1 (128)
HSGPA	85.0936	84.6576	85.2502
PUBLIC_PRIVATE	1.023	1.0217	1.0234
ENROLLED_AFTER_HS	Y	Y	Y
PREFERRED_COURSE	Y	N	Y
CET	44.592	44.4348	44.6484
GENDER	F	F	F
SCHOLAR	Y	Y	Y
FYGPA1	2.5133	2.6848	2.4517
FYGPA2	2.7963	3.1413	2.6723

Decision Support System

Figure 4. The Decision Support System Main Menu

Navigating the Decision Support System
To access the Decision Support System, the user goes to the homepage and then finds the link for the homepage of the DSS Main Menu. This homepage contains a link to the Decision Support System that contains four pull-down menus. The first menu has four types of information about the DSS itself including a description of the system, a glossary of terms and definitions, the technical specifications of the system, and answers to a list of frequently asked questions.

The College Completion module allows the user to examine student data individually, by course, department, and college and enables the user to choose 9 relevant attributes to run k-means clustering.

The Headcount and Characteristics is the third module and includes student demographic data by course, department, and college. This module allows the user to profile the students registered in a program by every data element included in the college application from demographics such as age, gender and civil status.

The What to Do module allows the administrator to choose from the list on what action to be taken based on the result of clustering. The options are: send message to student, schedule a dialogue with the student, notify the parents, advice student to consider other program.

DSS Evaluation

The system design has been carried out with ease of use being one of the highest criteria in the design of the software. It has a user-friendly Graphic User Interface (GUI). This allows the user to obtain needed information with a few mouse clicks.

The software has not yet entered widespread use. However, since the university is currently using a client-server version in its SARIAS the system can be easily added to its server. The advantage of the client-server version is that a central database can be maintained which will be updated centrally and all client user names and passwords can be controlled at the server side, hence allowing security and confidentiality.

The software is highly adaptable and expandable due to its modular nature of design. Additional functionality can be added and other campuses can easily adapt the software.

There is no cost to operate the software as it can be used by non-experts. The only cost foreseen is in the maintenance of the software. This will need the attention of an operator in order to manage access, privileges, user profiles and data update over the network.

V. **Conclusion**

The importance and wealth of information that can be extracted from student registration and grades databases has been shown.

It is thought that the software can be implemented in the university allowing sharing of information between administrative staff over the campus and local area networks automatically and in real-time without making any necessary data update.

The system will be reasonably cheap to maintain as it will have a client system that will require little support, and a server system that will have automatic data updating if the server is linked to the university MIS system. As student records databases are usually well established in registration systems, they will provide the most up-to-date data to the users.

The system has shown us the potential of academic data mining and the use of academic decision support software. We have however only used one data mining technique and several techniques can be used to analyze, interpret and use in the assessment of student success.

Our future work will be directed towards improving the data integration routines and enhancing the user interface to enable intuitive and flexible interactive visual exploration of the accumulated data and to use other data mining techniques other than clustering.

Acknowledgment

References

[1] E. Y. Angeles, National Higher Education Research Agenda 2 – NHERA 2 2009-2018. Commission on Higher Education, Philippines, 2009.

[2] V. P. Bresfelean and N. Ghisoiu, "Higher education decision making and decision support systems," Issue 2, Volume 7, February 2010, ISSN: 1790-1979.

[3] V.T.N. Chau and N. H. Phung, "A knowledge driven education decision support system," Computing and Communication Technologies, Research, Innovation and Vision for the Future. 2012 IEEE RIVF International Conference Ho Chi Minh City, 2012.

[4] D. Z. Deniz and I. Ersan, "An Academic Decision-Support System Based on Academic Performance Evaluation for Student and Program Assessment,". Int. J. Engng Ed. Vol. 18, No. 2, pp. 236±244, 2002

[5] T. Feghali, I. Zbib and S. Hallal, "A Web-based decision support tool for academic advising," Educational Technology & Society, 14(1), 82-94, 2011.

[6] S. B. Kotsiantis, "Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades," Artificial Intelligence Review. April 2012 Vol. 37, Issue 4 pp. 331–344, 2012 .

[7] H. M. Nagy, W. M. Aly and D. F. Hegazy, "An educational data mining system for advising higher education student," International Journal of Computer, Information and Engineering Vol. 7, No. 10 2013.

[8] O. J. Oyelade, O. O. Oladipupo and I. C. Obagbuwa, "Application of K-means clustering algorithm for prediction of students' academic performance," *(IJCSIS) International Journal of Computer Science and Information Security,* Vol. 7, _No. 1, 2010.

[9] A. M. Paz, B. D. Gerardo and B. T. Tanguilig. III, "Development of college completion model based on K-means clustering algorithm," International Journal of Computer and Communication Engineering, Vol. 3, No. 3, pp. 172-177, 2014.

[10] S. Vinnik and M. H. Scholl, "Decision support system for managing educational capacity utilization in university. International Conference on Engineering and Computer Education, 2005

About Authors:

Allen M. Paz is a faculty member at the Isabela State University. She is currently taking her Doctor in Information Technology at the Technological Institute of the Philippines, Quezon City. She is presently working on her research in Data Mining for her Dissertation.

Bobby D. Gerardo is currently the Vice President for Administration and Finance of West Visayas State University, Iloilo City, Philippines. He is a referee to international conferences and journal publications. He is interested in the following research fields: distributed systems, telematics systems, CORBA, data mining, web services, ubiquitous computing and mobile communications.

Bartolome T. Tanguilig III is currently the Assistant Vice President for Academic Affairs and concurrent Dean of the College of Information Technology Education and Graduate Programs of the Technological Institute of the Philippines, Quezon City.