Assisting Higher Education in Assessing, Predicting, and Managing Issues Related to Student Success: A Web-based Software using Data Mining and Quality Function Deployment

Amar Sahay, Ph.D*
Professor, Production and Supply Chain Mgt.
School of Business and Technology, Salt Lake Comm. College/University of Utah

Karun Mehta
Senior Research Engineer
QMS, LLP
Salt Lake City, UTAH

Abstract

This research seeks to develop a software system to assist higher education in assessing and predicting key issues related to student success. The software uses several data mining algorithms and quality tool such as, quality function deployment to study and predict issues including but not limited to enrollment management, dropout rate, time to degree, and suggest ways to improve courses and programs. Data mining knowledge discovery and predictive modeling tools along with quality function deployment (QFD) are used to uncover and understand hidden patterns in vast databases to understand student related issues and suggest ways to improve them. This research aims at: (1) discovering knowledge from massive data present in database of higher educational institution based on data mining, (2) using that knowledge for: classification, categorization, estimation, and visualization, (3) using predictive models including multiple regression, non-linear and logistic regression, to predict the critical student issues, (4) using the quality function deployment to address the voice of the students and other stakeholders and determine a prioritized list of critical factors that need to be deployed to design, develop, and improve the courses and programs vital to student success.

Key Words: Data Mining, Knowledge Discovery in Data Mining (KDD), Predictive Modeling, Quality Function Deployment (QFD), Analytical Hierarchal Process (AHP), Educational Data Mining (EDM),

Introduction

One of the biggest challenges that higher education faces today is predicting the academic paths of student. Many higher education systems specially, community colleges are unable to guide students in selecting career paths, deciding majors, and detecting student population who are likely to drop out because of lack of information and guidance from the school system. To better manage and serve the student population, the institutions need better assessment, analysis, and prediction tools to analyze and predict student related issues. These tools can be very helpful in managing and assisting students in community colleges as well as four year institutions that serve thousands of students in their various academic and vocational programs.

Data mining tools combined with Quality Function Deployment (QFD) can be effectively used in addressing and solving problems critical to student success. Data mining is the process of sampling, exploring, modifying, and assessing large amounts of data to uncover previously unknown patterns (SAS, 2005). It has been used successfully in discovering knowledge from large data bases (knowledge discovery in Data Mining or KDD) in different areas including science, business, engineering, finance, and fraud detection. One of the main objectives of data mining is to explore and model relationships in large amounts of data using predictive modeling. Quality Function Deployment (QFD) is a tool that has been successfully applied in quality engineering to enhance quality of products and processes, and reduce the cost of poor quality. This is a tool to address the 'voice of widely varying stakeholders' to prioritize the issues critical to student success or, in other words determining the list of critical factors in order of their importance to improve the process and achieve student success. In this research, the idea of QFD is extended to the field of education and is used to prioritize the issues critical to student success. The factors obtained using the data mining and predictive modeling process will be prioritized and deployed in order of importance to achieve the overall objectives of achieving student success. The QFD algorithm uses a correlation matrix and other mathematical technique to prioritize the critical issues by ranking these issues in order of importance. This helps the faculty and the department to deploy or act on the most critical issues in order of their importance. This tool also allows comparing an institution to the best in class for benchmarking. The QFD process will be designed as a part of overall software system. The application of the QFD process can be extended to the process of individual course and programs assessment and can be helpful in assessing classes, departments, and also other planning issues.

Critical Issues in Higher Education and Possible Solutions

One of the tasks important to an institution is to identify those students who are less likely to return from semester to semester. Having these students identified soon enough for the institution to tailor its interventions and marketing strategies will greatly enhance the persistence rate. *Although, various types of studies exist to analyze student persistence patterns, it still remains a challenging task to accurately predict a currently enrolled student's likelihood of returning to school the next term.* The tools and software proposed through this project can be very helpful in managing and assisting students in comprehensive community colleges in managing and improving their various academic and vocational programs. The software can be effectively used in other Utah institutions and two-year and four year institutions nationally.

To this date, a majority of the research on online student persistence has focused on finding a causal relationship between one variable and its effect on persistence. There is no research to date that has examined student persistence in online learning from a holistic approach, that is, one that takes into account fuzzy measurements such as student background characteristics along with measures of student behaviors and perceptions. To gain a clearer understanding of this phenomenon, it is important to develop a holistic and reliable system that tries to examine, assess and predicts, how multiple variables influence a student's decision to enroll or persist in a course, while taking into account the inherent vagueness and fuzziness of the factors involved. The predictive data mining model for managing student enrolment and retention should be holistic in order to effectively include the student characteristics and environmental factors which are fuzzy and difficult to express and quantify. This can be achieved by integrating advanced quality tools with fuzzy systems which can accurately translate the VoS (Voice of Students) into concrete outcomes by effectively mapping the student information to functional requirements.

For any academic institution, it is a challenging task to effectively capture, analyze and understand student's genuine requirements (VoS) by traditional methods. The student requirements are qualitative and tend to be imprecise and ambiguous due to their linguistic origins. Student attributes generally tend to be linguistic, qualitative and non-technical in nature and difficult to model via traditional methods. Variables which are used to describe student requirements are often improperly understood and expressed in abstract, fuzzy, or conceptual terms, which can culminate in vague assumptions and implicit inference. In order to capture student needs more effectively, the transformation process that converts student verbatim constructs, which are often tacit and subjective, into an explicit and objective statement of student needs, has to be emphasized. Poor understanding of student requirements and inaccurate assumption made during the elicitation and analysis of requirement information can have negative implications for academic institutions. Any predictions concerning student behavior and academic performance inferred from prediction models based on this information will be error-prone and misleading.

Background and Research Focus

This research focuses on potential applications of data mining, predictive modeling, and quality function deployment in higher education. Educational data mining (EDM) is "the process of uncovering hidden trends and patterns that lead them to predictive modeling using a combination of explicit knowledge base, sophisticated analytical skills and academic domain knowledge [Luan, Jing, 2002]." Data mining has been used successfully in science, engineering, business, and finance to extract previously unknown patterns in the data bases containing massive amount of data and to make predictions that are critical in decision making and improving the overall system performance. For example, in higher education, data mining can be used to predict the drop-out rate, transferability, retention, persistence, prediction of student population unlikely to return to school and other critical factors. These results allow the administration to act and intervene on time. In recent years, data mining is finding larger and wider applications in higher education. It is showing an increasing trend in institutional research. The reason for this is the growing interest in knowledge management and in moving from data to information and finally to knowledge discovery. A Tiered Knowledge Management Model (TKMM) in Figure 1 [Jing Luan, 2002] shows that data mining reflects the highest level of knowledge attainment that requires skills in data domain. The lowest level represents the data domain (Tier one), the next level is data querying and presentation (Tier two), and the top level represents mining, machine learning/artificial intelligence (Tier three). Data mining is at the top tier and depends on the lower tiers.

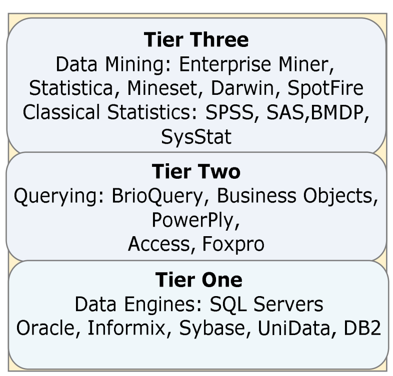

Figure 1: Knowledge Management Model

Educational data mining can be divided into following categories [Romero and Ventura, 2007]:

- Statistics and visualization
- Web mining
    - Clustering, classification, outlier detection
    - Association rule mining and sequential pattern mining
    - Text mining

According to Baker, the educational data mining has been classifies as:

- Prediction
    - Classification

    - Regression Analysis
    - Density Estimation
- Clustering
- Relationship mining
    - Association rule mining
    - Correlation mining
    - Sequential pattern mining
    - Casual data mining
- Discovery with modes

Historically, the most prominent category in educational data mining has been the relational mining.

In this research, we focus on knowledge discovery in data mining (KDD) and predictive modeling applications in education. This area is known as educational data mining (EDM). Education data mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the setting which they learn in [Baker, Yacef, 2009]. The data mining methods applied in educational research come from different areas including classical data mining, machine learning, psychometrics and statistics, informational visualization, and computational modeling.

Key Applications of Data Mining in Education and Current Trend

A number of applications have been reported where data mining has been applied in different areas. Some of the prominent research areas of educational data mining include computer supported collaborative learning, individual learning from education software, computer-aided testing, and the factors associated with student failure or non-retention in courses [Baker, Yacef, 2009]. Recently, there has been a shift in the research areas and also in the data mining methods being applied. Between 1995 and 2005, *relationship mining methods* was the most prominent area of data mining research in education. Approximately 43% of papers in those years involved relational mining methods. The second most prominent area in data mining research was *predictive data mining* with 28% of research papers involving prediction methods of different types. The third prominent research area in this period was human *judgment/exploratory data analysis* and *clustering* with 17% and 15% of research applications. In recent years (2008-2009) the trend and research in data mining educational application has shifted with *predictive data mining* dominating all research area representing 42% of all data mining research papers. The relationship mining methods that were dominant between1995-2005 moved to fifth place with only 9% of the research papers. The applications involving judgment/exploratory data analysis and clustering methods were approximately the same with 12% and 15% of the applications in 2008-09. A new method, significantly more prominent in recent years is *discovery with models* with 19% of all research reported in this area. [Barker, Barnes, and Beck 2008; Barnes Desmarias, Romero and Ventura, 2009]. Figure 2 shows the current trend and research focus in educational data mining.

Most reported research in Data Mining considered one or two issues related to higher education with specific data mining methods and none of these suggest how to deploy and implement critical issues and in what order the institutions implement the factors to correct the problems. There is a need to determine a priority list of the issues from the most important to least important issues that will guide higher institutions to deploy the most important and critical issues, intervene at the right time and take proper actions. In this research we integrate a quality function deployment to guide the institutions in the effort of deploying the critical factors.

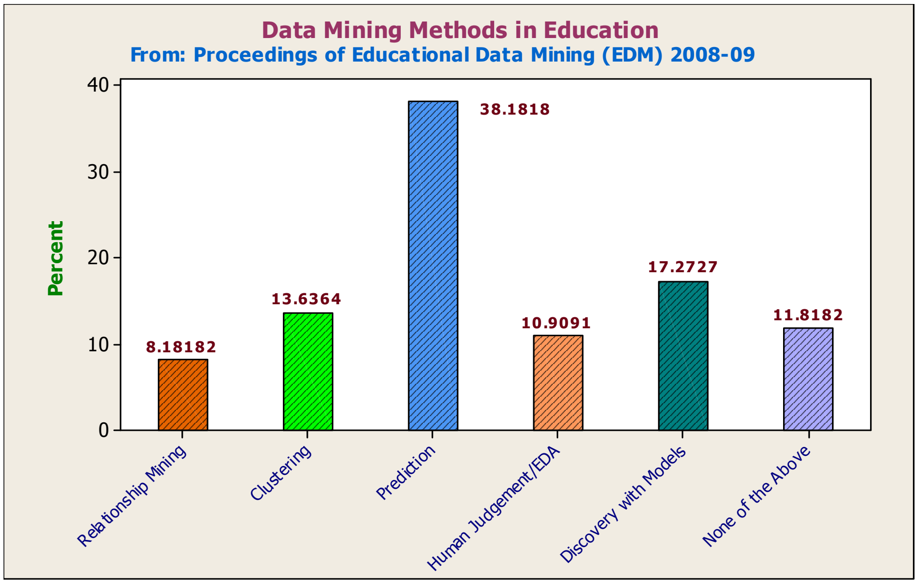

Figure 2: Data Mining Methods in Education: Current Trend

This tool can also allow an institution to compare itself to the best institutions in its class. This is achieved by integrating a quality function deployment tool to data mining. There is no system developed yet that integrates the knowledge discovery tools, predictive modeling and quality function deployment to address multiple issues facing the higher education today.

The software system we propose to develop provides a unifying framework for integrating data mining, predictive modeling, and QFD. The outcome of this research will be a software product that is highly scalable and can be used in researching and analyzing issues related to both two-year and four year institutions. The software design will be flexible enough to adapt to the changing needs of the institutions in addressing and solving issues related to student success. The proposed system address the current trend in the data mining, integrates the knowledge discovery in data mining (KDD), predictive modeling, and quality function deployment. Besides education, such a data mining software system has application in varied areas including finance, banking, manufacturing, and service industries.

Market for Data Mining:

The size of research estimate provided by Giga Research estimates that the size of data mining market has passed the billion dollar mark including software and services (according to consulting and service bureau). The conservative estimate of the data mining market size by other research organizations is between $200 and $700 million. However, the common agreement among various analysts suggests that data mining market represents the fastest growing business intelligent component. Data mining currently represents about 15% of the business intelligence market. The Educational Data Mining (EDM) is fairly new and one of the fastest growing areas of data mining and predictive modeling. It provides unique opportunities and institutional research. The software we intend to develop by integrating data mining, predictive modeling (discovery with models), and quality tools is unique and can be used in institutional research as well as in other business applications. In Utah, there is no study reported so far dealing with this type of research and product development. This tool will guide the higher education institution in studying the issues critical to student success.

Technical Discussion

Data mining is the process of discovering meaningful new correlations, patterns, and trends by shifting through large amounts of data stored in repositories and by using pattern recognition technologies as well as statistical and mathematical techniques [Gartner]. This research is an attempt to explore the data mining and quality techniques readily applicable in higher education. Our attempt is to link data mining, knowledge discovery in data mining (KDD), and quality function deployment towards a unifying framework for educational data mining application. Most data mining models applicable in current business and industry applications can also be applied in higher education research. This section will present a discussion on KDD process data mining and quality function deployment algorithms and show how these tools can be applied in solving problems faced by higher education today.

Guidelines to be followed: The most successful data mining projects follow the guidelines and the steps in the Cross-industry Standard Process for Data Mining (CRISP-DM). The CRISP-DM data mining methodology was initialized by three companies: SPSS, NCR, and Daimler Chrysler and was later sponsored by European Community research fund. We will follow the guidelines of CRISP-DM to build the educational application models. The CRISP-DM methodology can be described in terms of hierarchical process model consisting of tasks at four levels: phase, generic task, specialized task, and process instance. Phase is the topic level of the process. For example, the process and business understanding is the first phase of the project. Generic description of the task is data cleaning and data understanding. Specialized task would be the tasks within the generic task for example, data cleansing can consist of cleaning numeric values and cleaning categorical values. Process instance contains records of actions, decisions, and the result of actual data mining engagement. The phases of CRISP-DM are shown in Figure 3. The figure shows the life cycle of major phases of a data mining project. The project consists of six phases the sequence of which is not ordered.

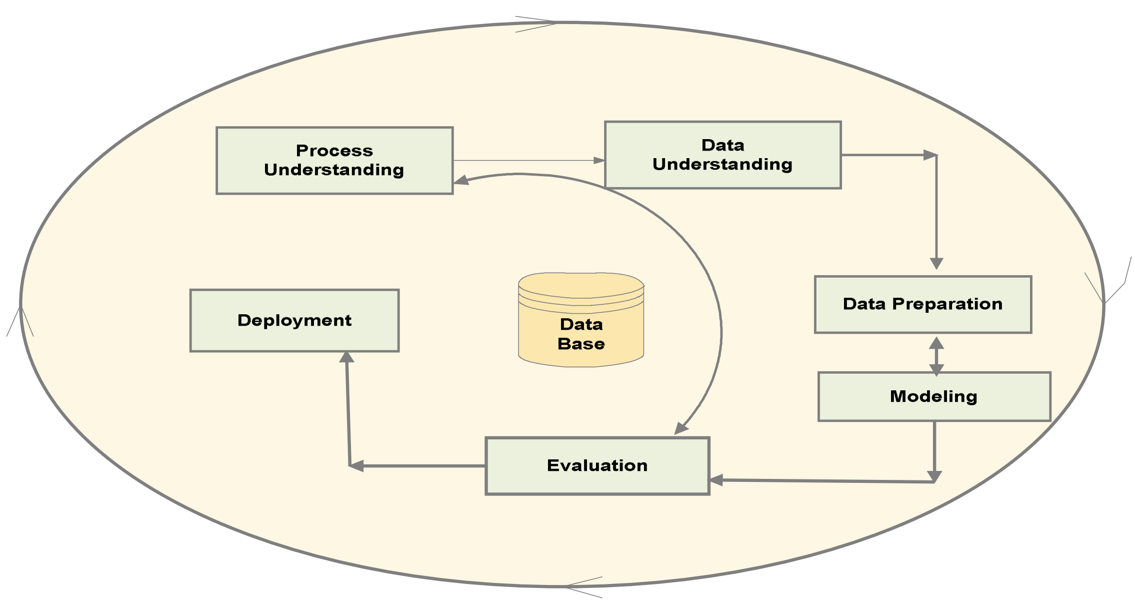

Figure 3: Phases of CRISP-DM Model

The six phases in Figure 3 provide the guidelines in designing the project. The first phase is the Business Processing Understanding which requires the statement of the business objective (what is to be achieved?). The Modeling phase involves selection of the models, conducting iterative tests on the data, and evaluating the model. For each step in the model there are specific tasks to be performed. The task of Data Preparation is critical that will require significant amount of time. The key to the success of data mining process lies in knowledge discovery in database (KDD).

As mentioned earlier, the main objective of this project is to study the issues related to student success in higher institutions. Many data mining tasks can be used for this. Before we discuss these tasks and details of the model, the data mining process and the knowledge discovery in data bases (KDD) needs an explanation.

Process of Data Mining and Knowledge Discovery in Databases (KDD)

A data mining process involves multiple stages, KDD involves the activities leading up to actual data analysis and evaluation and deployment of the results. The activities in KDD are shown in Figure 4 and described below.

- *Data Collection:* The goal of this phase is to extract the data relevant to data mining analysis. The data should be stored in a database where data analysis will be applied.
- *Data Cleaning and Preprocessing*: In this phase of KDD involves data cleansing and preparation of data to achieve the desired results. The purpose is to remove the noise and irrelevant information from the dataset.
- *Data Transformation*: This phase is aimed at converting the data suitable for processing and obtaining valid results. An example would be transforming a Boolean column type to integer.
- *Data Mining*: This purpose of data mining phase is to analyze the data using appropriate algorithm to discover meaningful patterns and rules to produce predictive models. This is the most important phase of KDD cycle.
- *Interpretation and Evaluation of Results:* In this final phase involves selecting the valid models for making useful decisions. All of the patterns determined from the previous data mining phase may not be meaningful. It is important to select the valid and meaningful patterns.

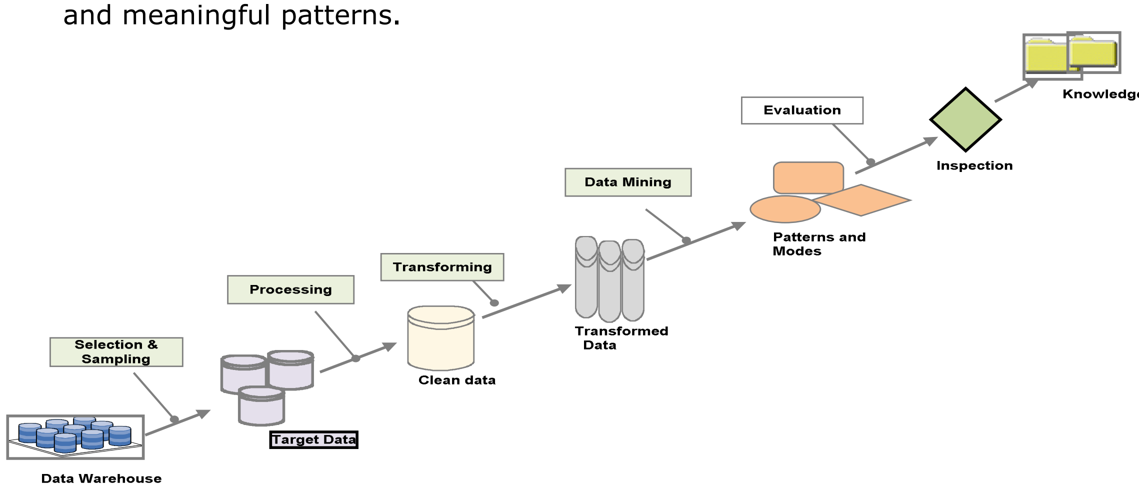

Figure 4: The Knowledge Discovery in Data Mining (KDD) Process

Data Mining Tasks, Algorithms, and Tools: Here we briefly describe some of the algorithms and predictive modeling tools we are going to use in this research. The most

common tools and their applications are described here. Data mining uses different algorithms that are diverse in their methods and have different objectives.
Tools: Several tools can be used to aid in the process preparing, processing and analyzing data. These tools may vary from EXCEL, to Access to simple SQL queries and visualization.

- *Data Visualization and exploration:* Data and results can be visualized using several graphical tools as well as through more specific tools such as symbolic data analysis which consists of creating groups by gathering individuals along one attribute. The aim of such analysis is to display data along certain attributes and make trends and clusters obvious.
- *Clustering:* Clustering algorithms aim at finding homogeneous groups in data. Several clustering methods including k-means clustering, hierarchic clustering are used to identify subgroups of cases based on a set of attributes.
- *Classification:* refers to assigning cases into different categories based on a predictable attribute. This is a widely used data mining task. Many problems in business and education involve classification.
- *Artificial Neural Network:* Uses a set of input variables, mathematical activation functions, and weighting of input variables to predict the value of the target variable(s). A neural network modifies its weight through iterative training cycle until the predicted output matched actual values. Once the network is trained, it can be used as a model that can be used against new data to predict different variables.
- *Rule Association:* Establishes cause-and-effect relationships and assigns probabilities or some type of certainty factors to support the conclusions. Rules are of the form "if <conditions>, then <conditions>" and are used to predict unknown variables.
- *Decision Trees:* A decision tree is a predictive modeling tool that is used for discovering new information within large databases or for building predictive models. Algorithms such as classification and regression trees (CART) or chi-square automatic induction (CHAID) are some of the common examples of decision trees.
- Besides the above predictive modeling tools and algorithms, there are other tools that will be needed to create models for the study.

Major Tasks and Proposed Solutions

The objectives, steps, and the proposed solution methods to achieve each of the objectives are outlined below.

(1) Determine the issues critical to student success in Utah institutions to be included in this research. For example, enrolment management, retention, or dropout rate are some of the issues to be addressed. If these factors are to be included in the study, obtain the higher education data that contain information related to those objectives. Also, create sample data that contain relevant information and also can be used as test data.
(2) Work with community colleges and other Utah institutions to determine and include the issues and challenges facing these institutions and use the data from these institutions to design, test, and validate the models and software system.
(3) Define data collection plan and procedures and collect data. The student and administration data are stored in many databases, data warehouses, and web servers across the higher education institution. The first step will be to pull the relevant data to a database where the data analysis will be performed. If needed, reduce the volume of the training dataset before extracting the pattern.
(4) Clean and transform the data and determine the data cleaning and transformation techniques to be applied. The data cleaning and transformation is the most resource-intensive step in data mining. The data collected needs to be cleaned and

transformed to remove the noise and irrelevant information out of the data set. The data transformation may be necessary to modify the source data into different formats in terms of data type and values. The data transformation techniques may require data type transformation, grouping, aggregation, missing value handling, detection and removal of outliers. Figure 4 above explains the data cleaning and transformation stage in the knowledge discovery in data mining (KDD) process.

(5) Create the Knowledge Based System for this study. Apply the data mining process to extract knowledge from the processed data in step 4. This data mining process will include the following steps: preprocessing data, applying data mining algorithms, and post processing the mined results. Figure 5 shows one example of knowledge discovery and creation of knowledge based system.

(6) Model Building. Build models to find the patterns. Once the data is cleaned, transformed, start the model building phase. Determine the appropriate models based on the goal of data mining project and the type of data mining task. Work with a domain expert to assist in the knowledge discovery and model building process. The model building is the core of this educational data mining project. Build multiple models using different algorithms and compare the accuracy of these models.

(7) Build Predictive Models. One of the most important goals of the study is to develop the predictive model to make predictions. Design trained models that can be used for prediction.

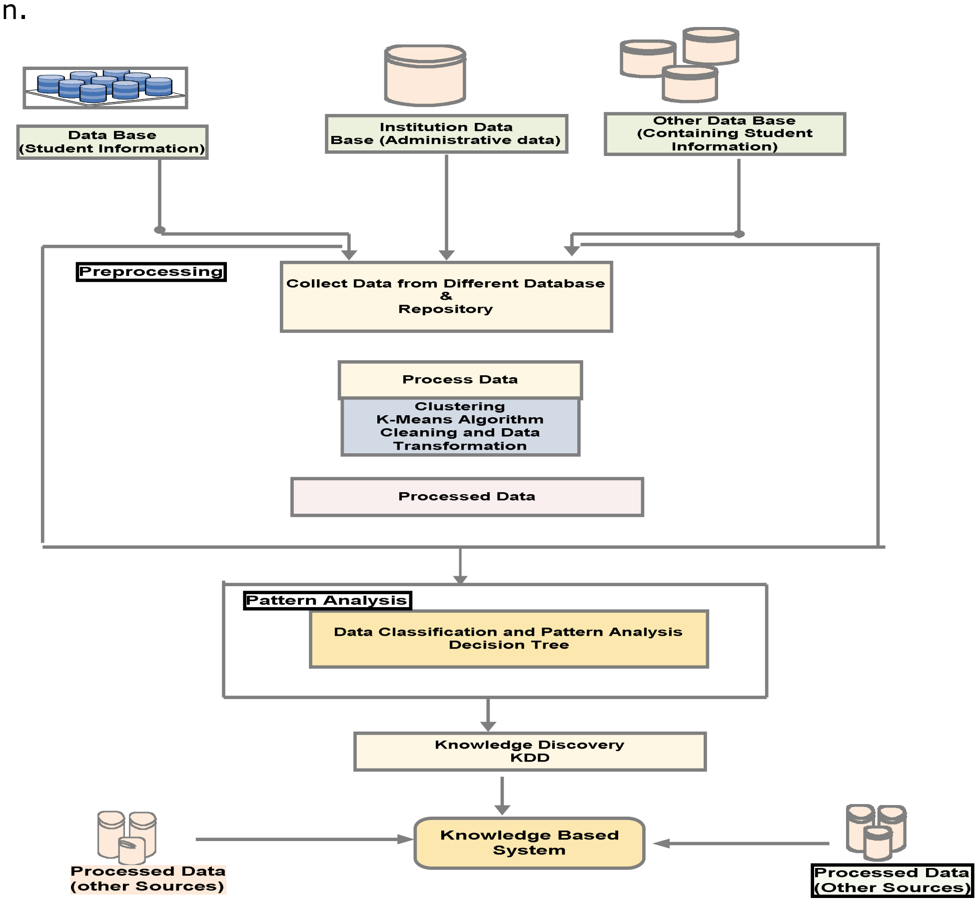

Figure 5: Example of Creating a Knowledge Based System

(8) Design the Quality Function Deployment (QFD) to prioritize the issues critical to student success. The critical factors obtained using the data mining and predictive modeling process are prioritized and deployed to achieve the overall objectives of

achieving student success. QFD a tool that has been successfully applied in quality engineering to enhance quality of products and processes, and reduce the cost of poor quality. This is a tool to address the 'voice of widely varying stakeholders' to prioritize the issues critical to student success or, in other words determining the list of critical factors in order of their importance to improve the process and achieve student success. Design the QFD as a part of overall software system. The application of the QFD process is extended to the process of individual course and programs assessment. The QFD process can be effectively used in assessing classes, departments, and also other planning issues. Figure 6 shows the structure of house of quality and how QFD can be applied in educational situation.

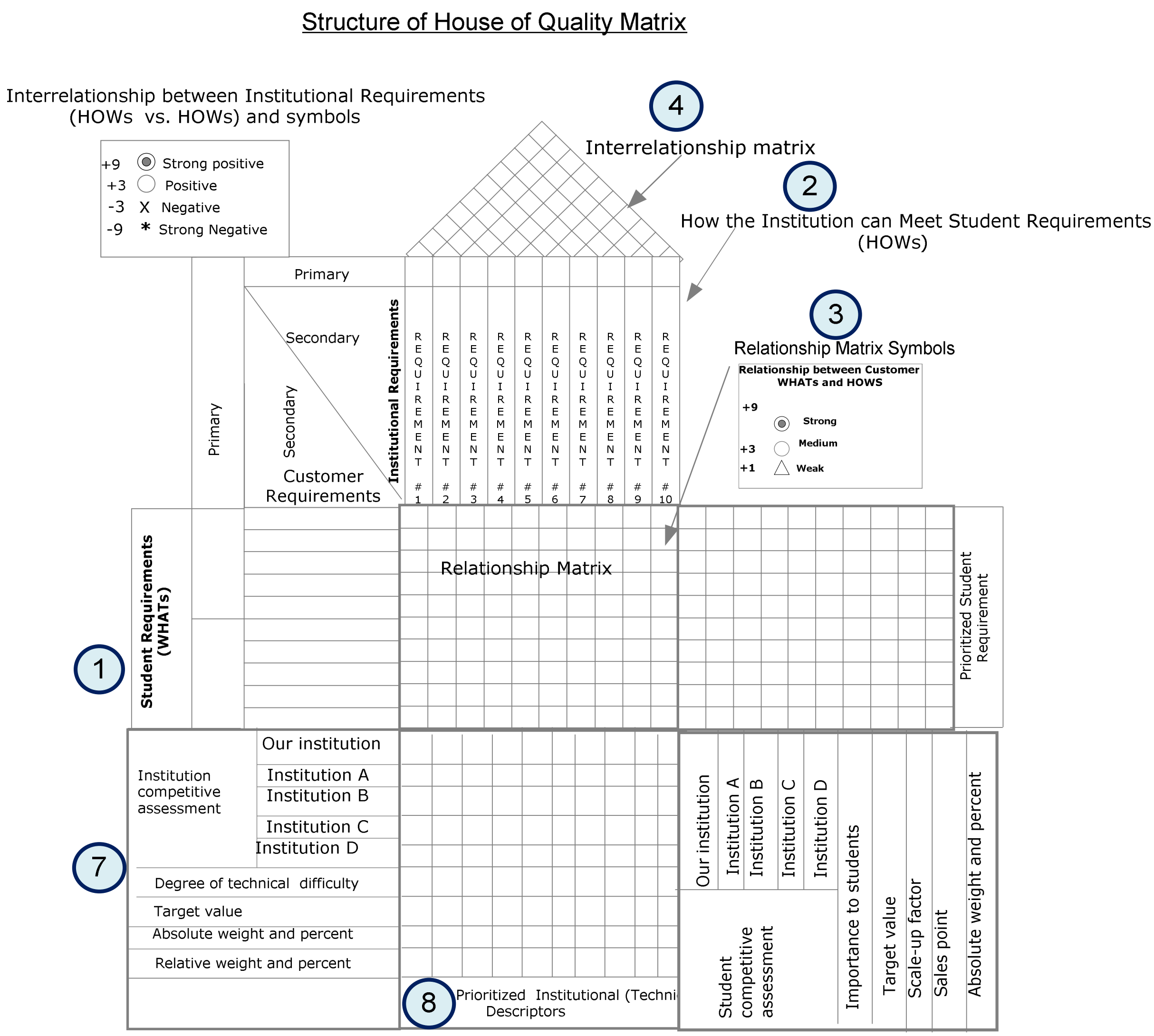

Figure 6: House of Quality and Quality Function Deployment (QFD) in Education

(9) Design the overall system architecture of the software system that integrates the knowledge discovery in data mining (KDD) process, predictive modeling, and quality function deployment.

(10) Create a prototype. Estimate the investment and commercial potential.

Summary

This research proposes a web-based software using Data Mining and Quality tools that will assist higher education in assessing, predicting, and managing issues related to student success. One of the biggest challenges that higher education faces today is predicting the academic paths of student. Many higher education systems specially, community colleges are unable to guide students in selecting career paths, deciding majors, and detecting student population who are likely to drop out because of lack of information and guidance from the school system. To better manage and serve the student population, the institutions need better assessment, analysis, and prediction tools to analyze and predict student related issues. These tools can be very helpful in managing and assisting students in comprehensive community colleges such as, Salt Lake Community College at Salt lake City, Utah that serves more than 60,000 students in their various academic and vocational programs. This paper suggested how to (1) study and predict the issues critical to student success at the higher educational institutions using the knowledge discovery and data mining tools, (2) use the data mining and predictive modeling tools such as, classification, clustering, categorization, estimation, visualization and others to extract knowledge from the databases of higher education and use predictive modeling tools to study and predict the variables such as enrollment, "student persistence," "dropout," "transfer," "retention," and "course success," (3) use quality function deployment (QFD), a tool to address the 'voice of widely varying stakeholders' to prioritize and deploy the issues critical to student success determined using the data mining and predictive modeling process, (4) design a flexible and interactive software system using the data mining, predictive modeling methods and algorithms, and quality function deployment that can be used in any higher education system to deal with the issues critical to student success. Although, data mining is widely used in Business, it has scarce applications in Education. Knowledge discovery using data mining and quality function deployment are relatively new areas for education and can be effectively used in addressing and solving problems critical to student success.

REFERENCES

Aksenova, S. S., Zhang, D. & Lu, M. (2006). Enrollment prediction through data mining. Information Reuse and Integration, 2006 IEEE International Conference

Bailey, B. (2006). Let the data talk: Developing models to explain IPEDS graduation rates. In J. Luan & C. Zhao (Eds.), New Direction for Institutional Research, no. 131. San Francisco: Jossey-Bass.

Baker, R.S.J.D. 2007. Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems. In *Proceedings of the ACM CHI 2007: Computer-Human Interaction conference*, 1059-1068.

Baker, R.S.J.D. in press. Data Mining For Education. In *International Encyclopedia of Education (3rd edition)*, B. MCGAW, PETERSON, P., BAKER Ed. Elsevier, Oxford, UK.

Baker R.S.J.D., BARNES, T. and BECK, J.E. 2008. 1 st International Conference on Educational Data Mining, Montreal, Quebec, Canada.

Barnes, T., Desmarais, M., Romero, C. and Ventura, S. 2009. Educational Data Mining 2009: 2nd International Conference on Educational Data Mining, Proceedings, Cordoba, Spain.

Bartneck, C. and Hu, J. 2009. Scientometric Analysis of the CHI Proceedings. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI2009)*, 699-708.

Beck, J. and Wolfe, B. 2000. High-level student modeling with machine learning. In *Proceedings of the International Conference on Intelligent tutoring systems*, 584-593.

Beck, J.E. 2007. Difficulties in inferring student knowledge from observations (and why you should care). *Proceedings of the AIED2007 Workshop on Educational Data Mining*, 21-30.

Bohannon, Tom, Baylor University Predictive Modeling in Higher Education, SAS Global Forum, 2007

Chang, L. (2006). Applying data mining to predict college admissions yield: A case study. In In J. Luan & C. Zhao (Eds.), New Direction for Institutional Research, no. 131. San Francisco:.

Dede, C. & Clarke, J. (2007). Data mining as an emerging means of assessing student learning. Presented at EDUCAUSE Annual Conference,.

Desai, Anoop and Jean-Claude Thomassian Engineering Course Design Based on Quality Function Deployment (QFD) Principles:Incorporation of Diverse Constituencies and Continuous Improvement, 38th ASEE/IEEE Frontiers in Education Conference October 22 – 25, 2008, Saratoga Springs, NY,

Fayd, Usama, Piatetsk, Gregory and Smyth, Padhraic , "*Knowledge Discovery and Data Mining:Towards a Unifying Framework,*" KDD-96 Proceedings.

Heathcote, L. & Dawson, S. (2005). Data mining for evaluation, benchmarking, and reflective practice in a LMS. Presented on E-Learn 2005: World conference on Elearning in corporate, government, healthcare, and higher education. Vancouver, Canada.

Herzog, S. (2006). Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. In J. Luan & C. Zhao (Eds.), New Direction for Institutional Research, no. 131. San Francisco.

Kelly, P. J. (2005). As America becomes more diverse: The impact of state higher education inequality. National Center for Higher Education Management System.

Luan, J. (2006). Using Academic Behavior Index (AB-Index) to Develop a Learner Typology for Managing Enrollment and Course Offerings: A Data Mining Approach. IR Applications, Volume 10.

Luan, J. & Zhao, C. (2006). Practicing data mining for enrollment management and beyond. In J. Luan & C. Zhao (Eds.), New Direction for Institutional Research, no. 131. San Francisco

Merceron,, A. and Yacef, K. 2003. A Web-based Tutoring Tool with MiningFacilities to Improve Learning and Teaching. In *11th International Conference onArtificial Intelligence in Education*

Piatetsky-Shapiro, G., Brachman, It., Khabaza, T.,Kloesgen, W., and Simoudis, E., 1996. An Overview of Issues in Developing Industrial Data Mining and KnowledgeDiscovery Applications, In Proceedings of KDD-96, Menlo Park, CA: AAAI Press.

Romera, C., Ventura, S., Espejo, P.G. and Hervas, C. 2008. Data Mining Algorithms to Classify Students. In *Proceedings of the 1st International Conference on Educational Data Mining*, 8-17.

Simoudis, E., Livezey, B., and Kerber, R. 1995. UsingRecon for Data Cleaning, In *Proceedings of KDD.95:First hdernational Conference on Knowledge Discoveryand Data Mining,* pp. 275-281, Menlo Park, *CA: AAAI*Press.

Sujitparapitaya, S. (2006). Considering student mobility in retention outcomes. In J. Luan & C. Zhao (Eds.), New Direction for Institutional Research, no. 131. San Francisco: Jossey-Bass.

Tan P., Steinbach M., Kumar V. Introduction to DATA MINING. Pearson Education,2006.

Tongshan, Chang, Ed.D. **Data Mining:** A Magic Technology for College Recruitment, a White paper

Zaine, O. 2001. Web usage mining for a better web-based learning environment. In *Proceedings of conference on advanced technology for education*, 60-64.

Zaine, O. 2002. Building a recommender agent for e-learning systems. In*Proceedings of the International Conference on Computers in Education*, 55–59