AC 2009-1669: APPLICATION OF EMERGING KNOWLEDGE-DISCOVERY METHODS IN ENGINEERING EDUCATION

Athanasios Tsalatsanis, University of South Florida

Athanasios Tsalatsanis received his Diploma degree in Production and Management Systems Engineering and his M.S. degree in Production Engineering from Technical University of Crete, Greece. He received his Ph.D. degree in Industrial Engineering from University of South Florida in 2008. Currently, he is a postdoctoral fellow with the Department of Industrial and Management Systems Engineering, University of South Florida. His research interests are in information systems, data analysis and decision sciences; control of autonomous intelligent systems; modeling analysis and control of discrete event dynamic systems.

Ali Yalcin, University of South Florida

Prof. Ali Yalcin received his B.S., M.S., and Ph.D. degrees in Industrial and Systems Engineering from Rutgers University, New Brunswick New Jersey in 1995, 1997 and 2000. He is currently an Associate Professor at the University of South Florida, Industrial and Management Systems Engineering Department, and an Associate Faculty member of the Center for Urban Transportation Research His research interests include modeling, analysis and control of discrete event systems, production planning and control, industrial information systems, data analysis and knowledge discovery, and engineering education research. He has taught courses in the areas of systems modeling and analysis, information systems design, production planning, facilities design, and systems simulation. He has publications in the areas of control of automated manufacturing systems, transportation systems, and autonomous vehicles, business process modeling, freight transportation systems analysis, people logistics, manufacturing and service information systems, and engineering education research. He also co-authored the 2006 Joint Publishers book-of-the-year textbook, Design of Industrial Information Systems, Elsevier.

Autar Kaw, University of South Florida

Autar K Kaw is a Professor of Mechanical Engineering and Jerome Krivanek Distinguished Teacher at the University of South Florida. He is the author of the textbook - Mechanics of Composite Materials, CRC-LLC Press. With major funding from National Science Foundation, he is developing award winning web-based resources for an undergraduate course in Numerical Methods. He is the recipient of the 2004 Council for Advancement and Support of Education (CASE) & the Carnegie Foundation for the Advancement of Teaching (CFAT) Florida Professor of the Year and the 2003 American Society of Engineering Education (ASEE) Archie Higdon Distinguished Mechanics Educator Award. His current scholarly interests include development of instructional technologies, integrating research in classroom, thermal stresses, computational mechanics, and mechanics of nonhomogeneous nanolayers.

APPLICATION OF EMERGING KNOWLEDGE DISCOVERY METHODS IN ENGINEERING EDUCATION

Abstract

The purpose of this study is to investigate the application of emerging knowledge discovery methodologies in analyzing student profiles to predict the performance of a student in a course. Knowledge discovery is the research area concerned with analyzing existing information and extracting implicit, previously unknown, hidden and potentially useful knowledge in an automated manner. The discovered knowledge is often represented by a set of rules or mathematical functions which has practical application. This type of knowledge can enable instructors to accommodate each student's learning needs and abilities as well as aid the students in appropriate course selection. In this paper we present a pilot study which demonstrates the analysis of student profiles from 60 students. The methodology used for knowledge discovery is based on Rough Set Theory which combines theories such as fuzzy sets, evidence theory and statistics. The results of the pilot study show that the knowledge discovery methodologies are likely to discover knowledge which may be overlooked using traditional statistical approaches. Our preliminary results indicate that knowledge discovery methodologies can be successfully used in predicting student performance. Based on the experiences gained from this work, specific future research directions and tasks to ensure a successful comprehensive implementation are discussed.

1. Introduction

Can we reliably predict the performance of a student in a particular course before he/she starts the course? Or can we recommend a specific set of course materials to certain students to improve their learning? What are the key factors that help answer these questions? Is it a student's past academic performance? Or is it their current work and/or class load? Or maybe it is their existing knowledge regarding the course material. More than likely it is an intricate combination of these and other factors some of which we do know and some we are yet to discover.

The overarching goal of our research is the development of a decision support system to enable both students and instructors to improve the quality of higher education. We envision that a decision support system which considers relevant information (including but not limited to a student's past performance in college, current work and course load, performance in the courses which are pre-requisites, etc.) can be designed to predict a student's performance in a course before the student begins the course. Based on this information, combined with the students learning style, a personalized learning strategy can be formulated to ensure the success of the student in the course.

We further envision that the development of such a decision support system can be achieved by utilizing emerging knowledge discovery methodologies (KDM). Knowledge discovery is the research area concerned with analyzing existing information and extracting implicit, previously unknown, hidden and potentially useful knowledge in an automated manner[1]. The discovered knowledge is often represented by a set of rules or mathematical functions which has practical application. Knowledge discovery methodologies have been used in a wide area of applications varying from agriculture[2,3] to total quality management[4,5] and healthcare[6,7].

Rough Set Theory (RST)[10] is an emerging knowledge discovery methodology that combines theories such as fuzzy sets[11], evidence theory[12] and statistics. RST is shown to be suitable for classification problems such as the prediction of student performance. The knowledge discovery process using RST is described in Figure 1.The fundamental source of data in the rough set framework is a two dimensional table. If the data table has missing values, it is preprocessed in several ways to complete it. The subsequent discretization step involves the representation of data using intervals and ranges in lieu of exact observations to define a coarser and more qualitative rather than quantitative representation from the data. The data mining step produces a set of if-then rules in two stages. First a minimal set of attributes are calculated and then the rules are generated based on this set of attributes. Finally, individual rules and patterns are ordered by a measure of "goodness" and inspected. The chosen rules are employed to classify new cases and measuring their performance. The advantages of RST over purely statistical methodologies are:

- Efficient algorithms for pattern extraction
- Consideration of quantitative and qualitative data
- Identification of attribute significance
- Easy to understand results
- Minimal set of decision rules

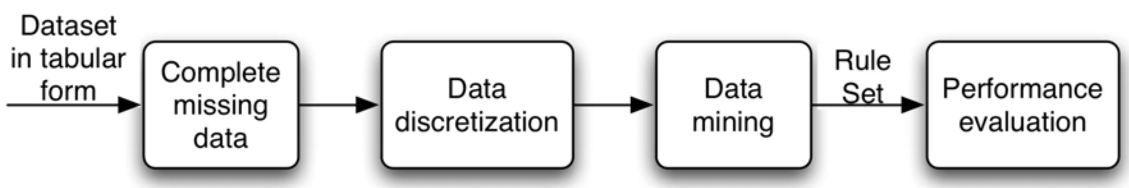

Figure 1. Knowledge discovery based on RST

In this paper, we present a knowledge discovery process based on RST using a pilot study on a small data set we have collected. It is important to note that we are not asserting that the results outlined in this paper are conclusive and/or valid, especially considering the small sample size and the limited range of relevant information upon which this study is conducted. Our main purpose here is to demonstrate the type of knowledge which may be gained from the application of knowledge discovery methodologies in education.

There is a clear need for further investigating the application of KDMs in education [8, 9]. Few approaches in literature utilize knowledge discovery methodologies in educational applications. Notable efforts include a data mining methodology to determine student desertion and retention[13], prediction of students' performance in an academic program[14], prediction of student performance based on grades in a prerequisite class[15], and mapping student learning mechanisms using KDM [16, 17].

The rest of the paper is organized as follows: Section 2 describes the details of this study. Section 3 presents the application of the knowledge discovery methodology on the student data; Section 4 discusses the results of the study; and Section 5 explores future research directions.

2. Description of the study

We define *student profile* as the set of attributes that capture information regarding the demographics, workload and student's past performance. These are candidate attributes which we believe to have a significant influence on the expected performance of the student in a particular course. The attributes considered are described in Table 1.

Table 1. Attribute definitions

Attribute	**Description**	**Definition**	**Attribute Range**
1	Age	The age of the student	<21: Less than 21 years old 22-26: Between 22 and 26 years old >26: Greater than 26 years old
2	Chld	If the student has children	Yes No
3	Crhr	The number of credit hours the student is taking during the semester	1-5: Between 1 and 5 Credit hours 6-11: Between 6 and 11 Credit hours >12: More than 12 Credit hours
4	Wrhr	The number of hours a student spends in work outside the school	0-10: Between 0 and 10 hours 11-20: Between 11 and 20 hours 21-30: Between 21 and 30 hours >30: More than 30 hours
5	Trnsf	If the student has been transferred from another institution	Yes No
6	Crch	If the student has made a career change	Yes No
7	Calc	The number of semesters elapsed since taking a prerequisite course	<4: Less than 4 semesters >4: More than 4 semesters
8	GPA	The overall GPA of a student	2.0-2.5 2.5-3.0 3.0-3.5 3.5-4.0
9	Pre-Test	The performance of the student in a test administered at the beginning of the class	Poor Fair Good

Our student population consists of 60 students registered in two offerings of the *Introduction to Linear Systems* course during the spring and summer terms in the 2008 academic year at a southeastern urban university. Information regarding student profiles was collected through a set of surveys and multiple choice examinations administered at the beginning of the course.

3. Application of RST to the data set

The analysis of the information captured from the student profiles was conducted based on Rough Set Theory. In RST, a data set is represented by a two dimensional table with n rows and m columns. Each row represents an object and each column represents a *condition attribute*. In this work, an object designates a student and a condition attribute designates an attribute included in the student profile. In most cases where a decision needs to be reached, an additional attribute, *decision attribute*, is incorporated in the data set. A system that encapsulates all objects, condition attributes and decision attributes is called a *decision system/table*.

Table 2 shows a part of the decision table used in this study. The attribute *Performance* is the decision attribute which indicates if a student has received a passing (A, B, C) or a failing grade (D, F) in the course.

Table 2. Decision Table

	Condition Attributes									**Decision attribute**
Student	**Age**	**Chld**	**Crhr**	**Wrhr**	**Trnsf**	**Crch**	**Calc**	**GPA**	**PreTest**	**Performance**
1	>26	YES	6-11	0-10	Yes	Yes	<4	3.5-4.0	Fair	Pass
2	22-26	NO	1-5	>30	No	No	>4	2.0-2.5	Good	Fail
3	<21	NO	1-5	>30	No	No	>4	3.0-3.5	Good	Pass
4	>26	YES	6-11	>30	Yes	Yes	>4	3.0-3.5	Fair	Pass

Knowledge discovery process entails several steps including data selection, data preprocessing and discretization before the actual data mining step. The surveys used to collect the data for this study were designed to provide the needed information in the desired format and therefore did not require these initial preparatory steps other than removing student profiles with incomplete/missing data.

3.1 Factors affecting student performance - Dependency of Attributes

A set of attributes *D* depends totally on a set of attributes *C* if all values of attributes from *D* can be uniquely determined by the values of attributes from *C*. The *degree of dependency* expresses the ratio of all elements of the universe which can be properly classified employing attributes of C. The concept of attribute dependency is helpful in determining which attributes or combinations of attributes are most significant in predicting student performance in the course. Table 3 lists the condition attributes used in our study and the degree of dependency of the decision attribute student performance. For example, using only the *Age* attribute we can correctly classify only 7 out of 60 students in this study. The decision attribute *Performance* has the greatest dependency with *GPA* among the condition attributes included in the study. Also notable is that the students performance on the pre-test given at the beginning of the semester was not significant in classifying student performance. However, due to reasons such as the small sample size and/or the design of the pre-test, further experimentation is required to generalize such findings.

Table 3. Dependency between attributes and Performance

Attribute	**Degree of dependency**
Age	7/60
Chld	5/60
Crhr	0
Wrhr	0
Trnsf	0
Crch	10/60
Calc	0
GPA	36/60
Pre-test	0

A *reduct* is the minimal subset of attributes that enables the same classification of objects of the universe as the complete set of attributes. The details associated with the algorithms to calculate reducts are fairly involved and will be omitted here. Table 4 shows the reducts of the complete data set. One of the important properties of reducts is the *core of attributes*. The core of attributes is the intersection of the attributes in the reducts. In a sense, the core is the most important set of attributes, since none of its elements can be removed without affecting the classification power of the attributes. Considering the reducts in Table 4, the core of attributes for our dataset is {Age, GPA}. The attributes {Wrhr, Calc.} appear in 4 of the 5 reducts indicating that these attributes are fairly important in classifying student performance in the course. Another important observation is that both the overall academic performance of the student (GPA) and the time that has elapsed since taking the pre-requisite calculus course (Calc.) are important attributes in classifying student performance. Our future studies will incorporate the grade that the student has received in the pre-requisite course as a condition attribute to investigate the dependency of student performance on this attribute.

Table 4. Reducts associated with the student profile dataset

{Age, Crhr, Wrhr, GPA}
{Age, Chld, Wrhr, Calc, GPA}
{Age, Wrhr, Calc, GPA, Initial CI}
{Age, Wrhr, Crch, Calc, GPA}
{Age, Crhr, Trnsf, Calc, GPA, Pre-Test}

3.2 Decision Rules

The language of *decision rules* are of the form *if* A *then* B $(A \rightarrow B)$, where A is called the *condition* and B the *decision* of the rule. Decision rules can be thought of as a formal language for drawing conclusions from data. In fact, every object in a decision system determines a decision rule and each reduct or set of condition attributes considered results in a different set of decision rules. For example using only the condition attribute GPA, the rules shown in Table 5 are extracted. The *RHS Support* indicates the number of students satisfying the condition of the rule and the *LHS Support* indicates the number of students satisfying the decision of the rule. Two probability factors, *certainty* and *coverage*, are associated with every decision rule. Certainty is defined by the conditional probability $P(B \mid A)$ which is the frequency of Bs in A, and coverage is defined by the probability $P(A \mid B)$ which is the frequency of As in B. A decision rule $(A \rightarrow B)$ is called "certain" if $P(A \mid B) = 1$ and "uncertain" otherwise.

Table 5.Decision rules based on GPA

Rule	**RHS Support**	**LHS Support (pass, fail)**	**Coverage**	**Certainty**
If GPA is 3.5-4.0 then Performance is Pass	17	17, 0	0.33, 0	1
If GPA is 3.0-3.5 then Performance is Pass	19	19, 0	0.37, 0	1
If GPA is 2.5-3.0 then Performance is Pass OR Fail	15	8, 7	0.16, 0.8	0.53, 0.47
If GPA is 2.0-3.5 then Performance is Pass OR Fail	9	7, 2	0.14, 0.2	0.78, 0.22

In Table 5, all students with GPA (3.0-3.5) and (3.5-4.0) can be certainly classified as passing the course. The sum of the number of students which can be certainly classified (36) divided by the total number of students (60) is the degree of dependency shown in Table 3. The remaining

24 students with GPA in the range (2.0-2.5) and (2.5-3.0) cannot be certainly classified without the use of other attributes as some have passed the course while others have failed. Table 6 shows the rules generated based on the *pre-test* attribute. Note that, none of the rules can be used to certainly classify student performance based only on this attribute.

Table 6. Decision Rules based on Pre-Test

Rule	**RHS Support**	**LHS Support (pass, fail)**	**Coverage (pass, fail)**	**Certainty (pass, fail)**
If Pre-Test is Fair then Performance is Pass OR Fail	29	26, 3	0.51, 0.33	0.90, 0.10
If Pre-Test is Good then Performance is Pass OR Fail	13	11, 2	0.22, 0.22	0.85, 0.15
If Pre-Test is Poor then Performance is Pass OR Fail	18	14, 4	0.27, 0.44	0.78, 0.22

Considering the set of reducts in Table 4, a total of 224 rules are generated which can be used to classify student performance. A portion of these rules with the highest RHS Support are listed in Table 7. The combination of attributes used in the condition side of the rule greatly reduces the number of students meeting these criteria leading to lower levels of support and coverage for the rule. On the other hand all rules are "certain" and unambiguously classify student performance.

Table 7. A subset of the decision rules used to classify student performance

Rule	**RHS Support**	**LHS Support (pass, fail)**	**Coverage**	**Certainty**
If Age is <21 AND Crhr is >12 AND Wrhr is 0-10 AND GPA is 3.5-4.0 the Performance is Pass	6	6, 0	0.12	1.0
If Age is <21 AND Wrhr is 0-10 AND Crch is NO AND Calc is <4 AND GPA is 3.5-4.0 then Performance is Pass	6	6, 0	0.12	1.0
If Age is <21 AND Chld is NO AND Wrhr is 0-10 AND Calc is <4 AND GPA is 3.5-4.0 then Performance is Pass	6	6, 0	0.12	1.0
If Age is <21 AND Wrhr is 11-20 AND Crch is NO AND Calc is <4 AND GPA is 3.0-3.5 then Performance is Pass	3	3, 0	0.06	1.0
If Age is <21 AND Chld is NO AND Wrhr is 21-30 AND Calc is <4 AND GPA is 3.5-4.0 the Performance is Pass	3	3, 0	0.06	1.0
If Age is 22-26 AND Chld is NO AND Wrhr is 21-30 AND Calc is >4 AND GPA is 3.0-3.5 then Performance is Pass	3	3, 0	0.06	1.0
If Age is <21 AND Crhr is >12 AND Trnsf is NO AND Calc is <4 AND GPA is 3.5-4.0 AND Pre-Test is Poor then Performance is Pass	3	3, 0	0.06	1.0
If Age is <21 AND Chld is NO AND Wrhr is 11-20 AND Calc is <4 AND GPA is 3.0-3.5 then Performance is Pass	3	3, 0	0.06	1.0
If Age is <21 AND Wrhr is 21-30 AND Crch is NO AND Calc is <4 AND GPA is 3.5-4.0 then Performance is Pass	3	3, 0	0.06	1.0

3.3 Classification and Performance

A KDM is implemented in two stages: *training* and *testing*. In the training stage the decision rules are generated and in the testing stage the decision rules are evaluated. In each of the stages a different dataset, namely *training dataset* and *testing dataset*, is used to avoid biases of the decision rules. To that end, the student profile dataset is randomly separated into two sub sets to be used for the training and testing stages. Each sub set contains information regarding 30 students. The reducts based on the training set of 30 students are shown in Table 8. Note that these reducts are different from those computed for the complete data set as shown in Table 4. Based on these reducts, 189 decision rules are generated.

Table 8. Reducts associated with the training dataset

{Age, Wrhr, GPA}
{Wrhr, Calc, GPA, Pre-Test}
{Age, Crhr, Trnsf, GPA, Pre-Test}
{Crhr, Wrhr, Trnsf, Calc, GPA}
{Chld, Crhr, Wrhr, Calc, GPA}
{Age, Trnsf, Calc, GPA, Pre-Test}
{Crhr, Trnsf, Calc, GPA, Pre-Test}

Using the decision rules generated in the training stage, each student of the testing dataset is classified into a specific class (pass or fail) based on a function called *degree of certainty* $\varphi(x)$. The degree of certainty is a normalized voting function describing the number of rules which classified a student (x) into a class divided by the total number of rules whose conditions describe the specific student. Formally:

$$\varphi(x) = \frac{votes(pass)}{votes(pass) + votes(fail)}$$

If the degree of certainty is greater than a threshold value $\tau \in (0,1)$ then the student is classified as passing student, otherwise as failing student. Usually, the threshold value is equal to 0.5.

The classification success of the KDM is evaluated using *confusion matrix* and *receiver operating characteristic* (ROC) analysis. The confusion matrix summarizes the ability of the derived rules to classify each student as failing or passing in terms of probabilities. The confusion matrix for the student performance study is shown in Table 9. The 189 rules generated based on the set of reducts shown in Table 8 were able to predict the performance of 3 out of 4 failing students and 21 out of 26 students correctly using a threshold of 0.5 as the degree of certainty.

Table 9. Confusion matrix

	Predicted			
actual		**fail**	**pass**	**Probabilities**
	fail	3	1	Specificity: 0.75
	pass	5	21	Sensitivity: 0.80
				Overall probability of success: 0.80

The definitions of the terms *sensitivity* and *specificity* in Table 9 are:

$$sensitivity = \Pr(prediction(pass) \mid actual(pass))$$
$$specificity = \Pr(prediction(fail) \mid actual(fail))$$

In other words, the generated rules were able to predict the performance of 80% of the students who passed the course as passing, and 75% of the students who failed the course as failing.

The ROC curve in Figure 2 represents how successful the set of rules are in classifying the students based on different values of the threshold, τ. Using different threshold values, different confusion matrices are computed. The values of specificity and sensitivity for each confusion matrix are used to create the ROC curve. The area below the ROC curve shows how well the decision rules classified the students into the correct class. If the area is less than 0.5 there is no classification capability. If the area is equal to 1, the classification is perfect. Regarding the student performance dataset, the area of the ROC curve depicted in Figure 2 is equal to 0.85. The area is computed using the trapezoidal approximation. Different decision rules may yield higher ROC curve area. The decision rules with the highest ROC curve area are the rules with the highest classification capability.

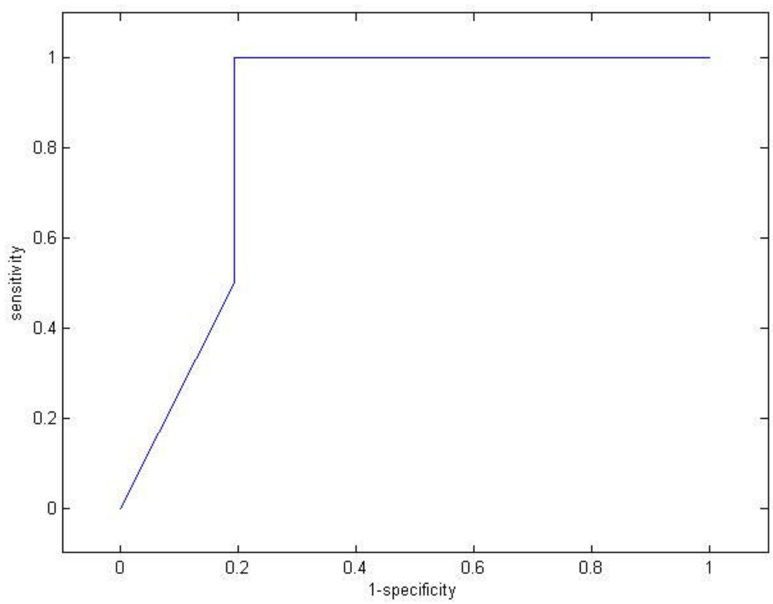

Figure 2. ROC curve

4. Discussion and Conclusions

As mentioned before, the knowledge discovered in this pilot study needs to be verified using larger data sets from multiple courses in multiple disciplines. However, our observations do warrant a brief discussion of the outcomes. The results in Table 3 show that the performance of a student in a course is dependent on the student's GPA, his/her age, career change and presence of children in their household. It is not surprising that all students with a GPA of 3.0 or above passed this course. On the other hand, it is interesting to observe that all students, over the age of 32, or with a child, or who had a career change also successfully completed the course. These three attributes indicate a significant dependency between the maturity of the student and their success in the course.

It is also interesting to observe that the student performance showed no dependency on the pre-test results, in fact the number of students who failed the course were almost evenly divided among the three ranges considered for pre-test results. The pre-test included 21 multiple choice questions mostly pertaining to students' knowledge of the material from pre-requisites and was styled as a concept inventory[18]. Naturally, the design of the test itself is a point of discussion. Nonetheless, the important questions to ask are: How significant is a student's knowledge of the course material in pre-requisite courses? Are students with a strong understanding of fundamental concepts more likely to perform better? While neither the results in this study, nor the design of the study are adequate to conclusively answer these questions, the knowledge discovery methods can provide interesting analysis approaches to answer such questions.

Considering the small sample size used in the study, the initial results pertaining to the prediction of student performance were surprisingly successful as shown by the confusion matrix and the ROC curve analysis. The real challenge is to be able to maintain or improve this level of success when the generated rules are used to predict student performance in multiple courses and in multiple disciplines which is the overarching goal of this research.

5. Future Research

One of the most important steps in the knowledge discovery process is data selection. In other words, the selection of candidate attributes which are believed to affect student performance in a course. Mining a set of attributes which may not be sufficiently focused not only increases the computational burden but may also result in irrelevant correlations. On the other hand, the strength of knowledge discovery lies in the fact that it can detect previously unknown and hidden correlations. Consequently, applying data mining techniques to a very limited data set will not utilize the full potential of this approach. Significant effort must be invested in determining an appropriate set of attributes by thoroughly examining the existing literature and relying upon experts' opinions.

Another important consideration is the design of the surveys used for data collection. In our pilot study student data was collected using arbitrary ranges. For example, the working hours attribute was collected using ranges 0-10, 11-20, 21-30 and 31 or more. RST offers several tools to discretize exact quantitative data[19] and discover ranges which may be more appropriate for knowledge discovery. Future work will also focus on the design of these surveys to collect data in a manner that can be best utilized by the selected data mining methodology.

Finally, there is significant work remaining in the design of surveys and examinations to determine students' learning styles as well as their knowledge of the course material covered in pre-requisite courses. This information is critical in formulating a personalized learning strategy for the individual student to ensure successful completion of the course.

Ultimately, from an implementation and usability perspective, several points must be taken into consideration. First and foremost, all data collection and analysis must be done automatically online so that a student can independently and privately utilize the envisioned decision support system. The course instructor's involvement must be limited to the design of the exams for data collection to minimize the time commitment to use such a system for his/her courses. Finally, the design of the system must be sufficiently flexible to be used in a wide range of courses, disciplines and institutions.

Bibliography

1. Polkowski, L., Tsumoto, S., Lin, T.Y., Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems, Physica-Verlag, New York, 2000.
2. Lee, S.W., Lerschberg, L., A methodology and life cycle model for data mining and knowledge discovery in precision agriculture, IEEE International Conference on Systems, Man and Cybernatics, vol. 3, pp. 2882-2887, 1998.
3. Ahmad, F., Zakaria, N.H., Osman,S.W.R., Transforming Information-Based Agricultural Portal to Knowledge-Based Agricultural Hub, IEEE International Conference on Information and Communication Technologies, pp. 1-4, 2008.
4. Wang, X. The Realization of Knowledge Discovery in Total Quality Management System, IEEE International Conference on Fuzzy systems and Knowledge Discovery, pp. 643-646, 2008.
5. Wang, K., Tong, S., Eynard, B., Roucoules, L., Matta, N., Application of Data Mining in Manufacturing Quality Data, International conference on Wireless Communications, Networking and Mobile Computing, pp. 5382-5385, 2007.
6. Berler, A., Pavlopoulos, S, Koutsouris, D., using key performance indicators as knowledge-management tools at a regional health-care authority level, IEEE Transactions on Information Technology, vol. 9, Issue 2, pp. 184-192, 2005.
7. Potamias, G.A., Moustakis, V.S., Knowledge discovery from distributed clinical data sources: the era for internet-based epidemiology, International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 4, pp. 3638-3641, 2001.
8. Masullo, M.J., Discovering the Academic Potential of Our Children, Golden Mean Column of the On-Line Decision Support Journal, available at: www.inivovision.com/library/R-9810.pdf
9. Luan, J., Data Mining As Driven by Knowledge Management in Higher Education – Persistence Clustering and Prediction, Keynote for SPSS Public Conference, UCSF, 2001.
10. Pawlak, Z., Rough set approach to knowledge-based decision support, European Journal of Operational Research, pp. 1-10, 1997.
11. Klir, G.J., Yuan, B., Fuzzy sets and fuzzy logic: theory and applications, Prentice Hall PTR, N.J., 1995.
12. Glenn, S., A Mathematical Theory of Evidence, Princeton University Press, 1976.
13. Salazar, A., Gosalbez, J., Bosch, I., Miralles, R., Vergara, L., A case study of knowledge discovery on academic achievement, student desertion and student retention, IEEE International Conference on Information Technology: Research and Education, pp. 150-154, 2004.
14. Ramasubramanian, P., Iyakutti, K, Sureshkumar, V., Thangavelu, P., Mining Analysis of SIS database using Rough Set Theory, International Conference on Computational Intelligence and Multimedia Applications, pp. 81-87, 2007.
15. Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer,G., Punch, W.F., Predicting Student Performance: An Application of Data Mining Methods With the Educational Web-Based System LON-CAPA, Annual Frontiers in Education, pp. T2A-13-18, 2003.
16. Pimentel, E.P., Omar, N., An Architecture of a Computer Learning Environment for Mapping the Student's Knowledge Level, Issues in Informing Science and Information Technology, vol. 4, pp. 313-326, 2007.
17. Ahmad, N.B.H., Shamsuddin, S.M.H., Mapping Student Learning characteristics Into Integrated Felder Silverman Learning Style Using Rough Set Classifier, Postgraduate Annual Research Seminar 2007.

18. Hestenes, D., Wells, M., and Swackhamer, G., "Force Concept Inventory". The Physics Teacher, vol. 30, issue 3, pp 141-151, 1992.
19. Dai, J.H., Li, Y.X., Study on discretization based on rough set theory, IEEE International conference on Machine Learning and Cybernetics, vol. 3, pp. 1371-1273, 2002.