



EMPIRICAL STUDY OF DATA MINING TECHNIQUES IN EDUCATION SYSTEM

V. Sarala¹ Dr. V.V.Jaya Rama Krishnaiah

¹Asst. Professor, DNR College, Bhimavaram, saralavedala@yahoo.com

² Associate Professor, A.S.N. Degree College, Tenali

ABSTRACT

Educational institutions are important parts of our society and playing a vital role for growth and development of nation. In earlier days the information flow in education field was relatively simple and the application of technology was limited. However, as we progress into a more integrated world where technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Today, one of the biggest challenges that educational institutions face the explosive growth of educational data and to use this data to improve the quality of managerial decisions. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets. In this paper the applications of data mining in educational institution to extract useful information from the huge data sets and providing analytical tool to view and use this information for decision making processes by taking real life examples.

INTRODUCTION

In modern world a huge amount of data is available which can be used effectively to produce vital information. The information achieved can be used in the field of Medical science, Education, Business, Agriculture and so on. As huge amount of data is being collected and stored in the databases. Data Mining or data or knowledge discovery has become the area of growing significance because it helps in analyzing data from different perspectives and summarizing it into useful information. Data Mining can be defined as the process involved in extracting interesting, interpretable, useful and novel information from data. There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data from educational environments. The data can be collected from various educational institutes that reside in their databases. The data can be personal or academic which can be used to understand students' behavior, to assist instructors, to improve teaching, to evaluate and improve e-learning systems, to improve curriculums and many other benefits. Educational data mining uses many techniques such as decision trees, neural networks, support vector machines and many others. Using these techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for organization of syllabus, prediction regarding enrolment of students in a particular programme, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students and so on.

RELATED WORK

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes.

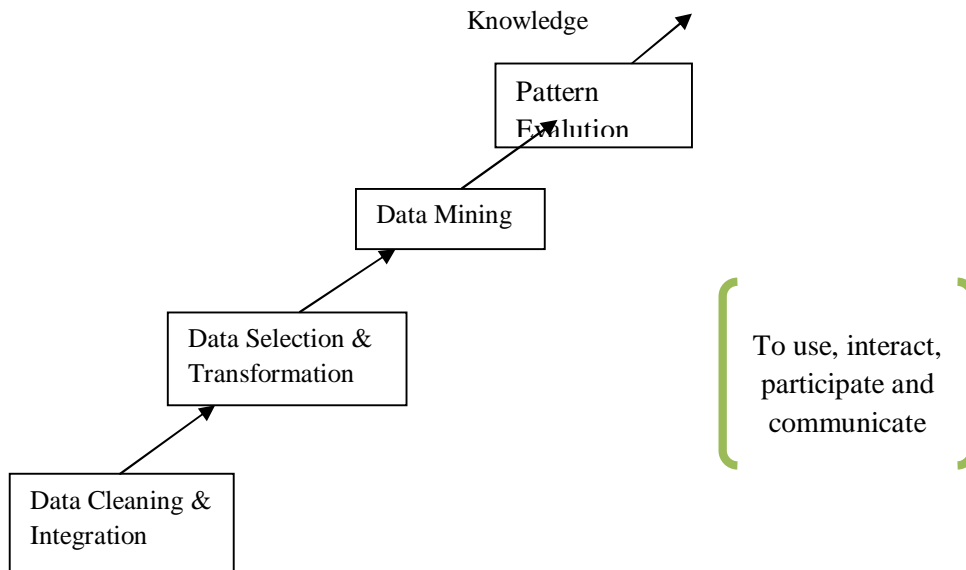
- 1) gave case study of using educational data mining in Moodle course management system. They have described how different data mining techniques can be used in order to improve the course and the students' learning. All these techniques can be applied separately in a same system or together in a hybrid system.
- 2) have a survey on educational data mining between 1995 and 2005. They have compared the Traditional Classroom teaching with the Web based Educational System. Also they have discussed the use of Web Mining techniques in Education systems.
- 3) discuss how data mining can help to improve an education system by enabling better understanding of the students. The extra information can help the teachers to manage their classes better and to provide proactive feedback to the students.
- 4) have described the use of data mining techniques to predict the strongly related subject in a course curricula. This information can further be used to improve the syllabi of any course in any educational institute.

5) describes how data mining techniques can be used to determine. The student learning result evaluation system is an essential tool and approach for monitoring and controlling the learning quality. From the perspective of data analysis, this paper conducts a research on student learning result based on data mining.

DATA MINING DEFINITION AND TECHNIQUES

Data mining refers to extracting or “mining” knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships in decision making.

The Steps of extracting knowledge from data



Association analysis:

Association analysis is the discovery of association rules attribute-value conditions that occur frequently together in a given set of data. Association analysis is widely used for market basket or transaction data analysis. More formally, association rules are of the form $X \Rightarrow Y$, i.e., “ $A_1 \wedge \dots \wedge A_m \rightarrow B_1 \wedge \dots \wedge B_n$ ”, where A_i (for $i \in \{1, \dots, m\}$) and B_j (for $j \in \{1, \dots, n\}$) are attribute-value pairs. Association rule $X \Rightarrow Y$ is interpreted as database tuples that satisfy the conditions in X are also likely to satisfy the conditions in Y ”.

Classification and Prediction:

Classification is the process of finding a set of models or functions which describe and distinguish data classes or concepts. The derived model may be represented in various forms, such as classification(IF-THEN) rules, decision trees, mathematical formulae, or neural networks. Classification can be used for predicting the class label of data objects. IF-THEN rules are specified as IF condition THEN

e.g. IF age=youth and student=yes then buys_computer=yes

Cluster Analysis:

Unlike Classification and prediction, which analyze class-labeled data objects, ‘clustering’ analyzes data objects without consulting a known class label. Clustering can be used to generate such labels. Clusters of objects are

formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Application of clustering in education can help institutes group individual student into classes of similar behavior. Partition the students into clusters, so that students within a cluster(e.g. Average) are similar to each other while dissimilar to students in other clusters(e.g. Intelligent, Weak).

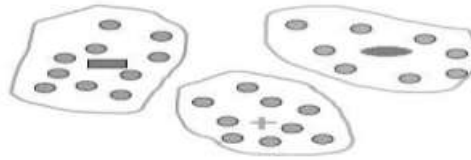
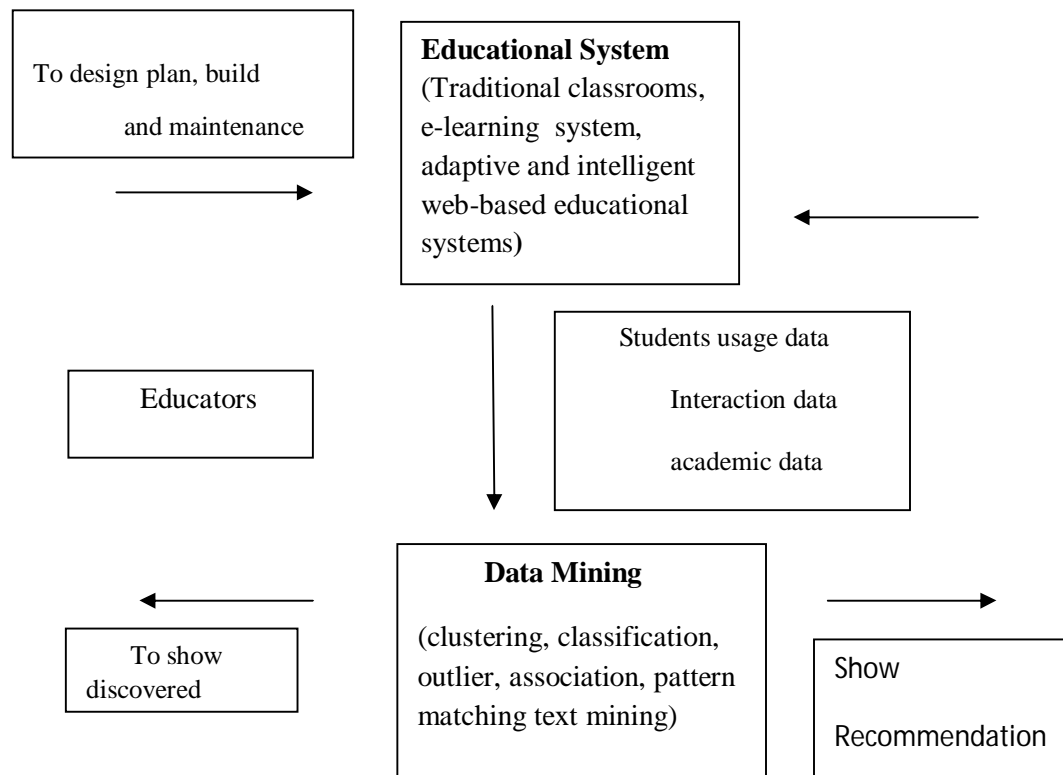


Fig: Picture showing the partition of students in clusters

The cycle of applying data mining in education system



The above figure illustrates how the data from the traditional classrooms and web based educational systems can be used to extract knowledge by applying data mining techniques which further helps the educators and students to make decisions.

Educational Tasks and Data Mining Techniques

There are many applications or tasks in educational environment that have been resolved through Data Mining. For example, Baker suggests four key areas of application for Educational Data Mining: Improving student models Improving domain models. Studying the pedagogical support provided by learning software. Scientific research into learning and learners.

And five approaches/methods:

Prediction, Clustering, Relationship mining, Distillation of data for human judgement, Discovery with models.

Educational Data Mining subjects/tasks:

Applications dealing with the assessment of the student’s learning performance. Applications that provide course adaptation and learning recommendations based on the student’s learning behavior. Approaches dealing with the evaluation of learning material and educational web-based courses. Applications that involve feedback to both teacher and students in e-learning courses. Development for detection of atypical students’ learning behaviors.

A. Organization of Syllabus:

It is important for educational institutes to maintain a high quality educational programme which will improve the student’s learning process and will help the institute to optimize the use of resources. A typical student at the university level completes a number of courses or subjects prior to graduation. Presently, organization of syllabi is influenced by many factors such as affiliated, competing or collaborating programmes of universities, availability of lecturers, expert judgments and experience. One of the application of data mining is to identify related subjects in syllabi of educational programmes in a large educational institute. In the association rule mining is used to identify possibly related two subject combinations in the syllabi which also reduce our search space.

For this purpose following methodology was followed to:

- Identify the possible related subjects.
- Determine the strength of their relationships and determine strongly related subjects.

THE SUBJECTS CHOSEN BY STUDENTS

Student id	Subject 1	Subject 2	Subject 3
1	Databases	Advanced Databases	Data mining
2	Databases	Advanced Databases	Data mining
3	Databases	Advanced Databases	Data mining
4	Databases	Advanced Databases	Visual Basic
5	Databases	Advanced Databases	Web Designing

Association Rules that can be derived from the above table are of the form:

$$(X, \text{subject 1}) \Rightarrow (X, \text{subject 2})$$

$$(X, \text{subject 1}) \wedge (X, \text{subject 2}) \Rightarrow (X, \text{subject 3})$$

$$(X, \text{“Databases”}) \Rightarrow (X, \text{“Advanced Databases”})$$

$$(X, \text{"Databases"}) \wedge (X, \text{"Advanced Databases"}) \Rightarrow (X, \text{"Data Mining"})$$

B. Predicting the Registration of Students in an Educational Programme:

Now a days educational organization are getting string competition from other Academic competitors. To have and edge over other organizations, needs deep and enough knowledge for a better assessment , evaluation, planning, and decision making. Data Mining helps organizations to identify the hidden patterns in databases; the extracted patters are than used to build data mining models, and hence can be predict performance and behavior with high accuracy.

For example, to efficiently assign resources with an accurate estimate of how many male or female will register in a particular program by using Prediction technique.

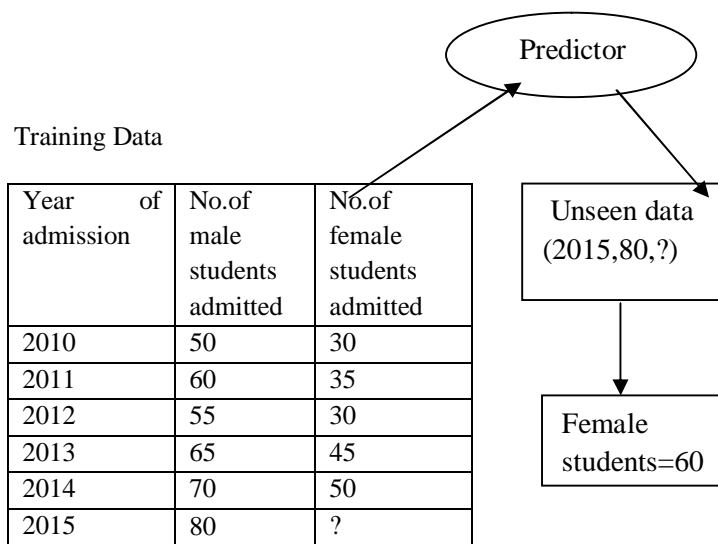


Figure: Prediction of female students in the coming year

We can use student participation data as part of the class grading policy. An instructor can assess the quality of student by conducting an online discussion among a group of students and use the possible indicators such as the time difference between posts, frequency distribution of the postings, duration between postings and replies etc. Given this data, we can apply classification algorithms to classify the students into possible levels of quality.

C. Predicting Student Performance:

It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling. The main objective of this paper is to use data mining methodologies to study students’ performance in the courses. Data mining provides many tasks that could be used to study the student performance. By this task we extract knowledge that describes students’ performance in end semester examination. The classification task is used to evaluate student’s performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information’s like

Attendance, Class test, Seminar and Assignment marks were collected from the student’s management system, to predict the performance at the end of the semester.

STUDENT RELATED VARIABLES

Variable	Description	Possible Values
PSM	Previous Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36% }
CTG	Class Test Grade	{Poor , Average, Good}
SEM	Seminar Performance	{Poor , Average, Good}
ASS	Assignment	{ Yes, No }
ATT	Attendance	{Poor , Average, Good}
LW	Lab Work	{ Yes, No }
ESM	End Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36% }

For example, PSM has the highest gain, therefore it is used as the root node as shown in figure.

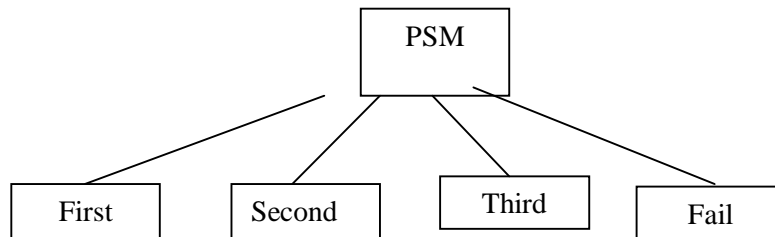


Figure: PSM as root node

D. Detecting Cheating in Online Examination:

We can say that online assessments are useful to evaluate students’ knowledge; they are used around the world in school education to higher education institutions. Now a days exams are conducted online remotely through the Internet and if a fraud occurs then one of the basic problems. Cheating is not only done by students but the recent scandals in business and journalism show that it has become a common practice. Data Mining techniques can propose models which can help organizations to detect and to prevent cheats in online assessments. The models generated use data comprising to different student’s personalities, stress situations generated by online assessments, and common practices used by students to cheat to obtain a better grade on these exams.

E. Identifying Abnormal / Erroneous Values:

The data stored in a database may reflect outlier/ noise, exceptional cases, or incomplete data objects. These objects may confuse the analysis process, causing over fitting of the data to the knowledge model constructed. As a result, the accuracy of the discovered patterns can be poor. One of the applications of Outlier Analysis can be detect the abnormal values in the result sheet of the students. This may be due many factors like a software fault, data entry operator negligence or an extraordinary performance of the student in a particular subject.

The result of students in four subjects

Student Roll No.	Marks in Subject1	Marks in Subject2	Marks in Subject3	Marks in Subject4
1	30	35	45	30
2	67	76	78	67
3	89	90	78	77
4	30	75	77	76
5	30	35	45	99

In the above shown table the result of the student in subject4 with roll no 5 will be detected as an exceptional case and can be further analyzed for the cause.

CONCLUSION:

In the present study, we have discussed the various data mining techniques which can support education system via generating strategic information. Since the application of data mining brings a lot of advantages in higher learning institution, it is recommended to apply these techniques in the areas like optimization of resources, prediction of students performance. The classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information like Attendance, Class test, Seminar and Assignment marks were collected from the student’s previous database, to predict the performance at the end of the semester.

REFERENCES

[1] C. Romero, S. Ventura "Educational data Mining: A Survey from 1995 to 2005", Expert Systems with Applications (33), pp. 135-146, 2007

[2] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, “Data Mining Model for Higher Education System”, European Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010

[3] K. H. Rashan, Anushka Peiris, “Data Mining Applications in the Education Sector”, MSIT, Carnegie Mellon University, retrieved on 28/01/2011

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

[5] Kumar, V. (2011). An Empirical Study of the Applications of Data Mining Techniques in Higher Education. *IJACSA - International Journal of Advanced Computer Science and Applications*, 2(3), 80-84. Retrieved from <http://ijacsa.thesai.org>.

[6] Manoj Bala, Dr.D.Bojha , “Study of Applications of Data Mining Techniques in Education”. *IJRST – International Journal of Research in Science And Technology*.