Way to Learning Dropouts in Distance Education -A Study using Data mining Techniques

B Deepthi Chary[1], Y Sucharitha[2] , P Chandra shaker Reddy[3]
[1]*Asst. Professor, Dept of CSE, Malla Reddy Inst. of Engg & Tech ,AP,INDIA, deepthi.chary@gmail.com*
[2]*Asst. Professor, Dept of CSE, Malla Reddy Inst. of Engg & Tech ,AP,INDIA,suchi.yadala@gmail.com*
[3]*Asst. Professor, Dept of CSE, Malla Reddy Engg. College,AP,INDIA,chandu.pundru@gmail.com*

ABSTRACT:

Student dropout occurs quite often in universities providing distance education and the dropout rates are definitely higher than those in conventional universities. Limiting dropout is essential in university-level distance learning and therefore the ability to predict students, dropout could be useful in a great number of different ways. Distance education meets the needs of those students might otherwise be unable to attend on-campus classes, due to distance time constraints. One of the biggest benefit is the issue of flexibility and time. Because students are confined to a classroom for a certain number of hours on a given day , they can approach their coursework with flexibility and complete lessons when it suits their schedule. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from large data sets. This paper focuses the on data mining in educational institution to extract useful information from the huge data sets on e-learning systems and environments for distance learning for such dropout students.

Keywords – *Data Mining, distance learning Drop-outs, prediction, coursework*

I .INTRODUCTION

Today, 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, traditional statistical techniques and database management tools are no longer adequate for analyzing this huge amount of data. Data Mining (sometimes called 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. 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 originating from educational environments.

The data can be collected from various distance learning educational institutes that reside in their databases. The data can be personal or academic which can be used to understand students' behavior (socio economic factors for dropping out), to assist instructors, to improve teaching, to evaluate and improve e-learning systems, to improve their coursework.

The objective is to identify the potential areas in distance education where data mining techniques can be applied in the field of Higher Education and e-learning activities by improving their knowledge skill development through continuing education Programs to meet their needs.

II. DATA MINING DEFINITION AND TECHNIQUES

Data mining (also known as Knowledge Discovery in Databases - KDD) has been defined as "The nontrivial extraction of implicit, previously unknown, and potentially useful information from data". Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move Beyond the analyses of past events provided by retrospective tools typical of decision support systems. The sequences of steps identified in extracting knowledge from data are: shown in Figure Knowledge
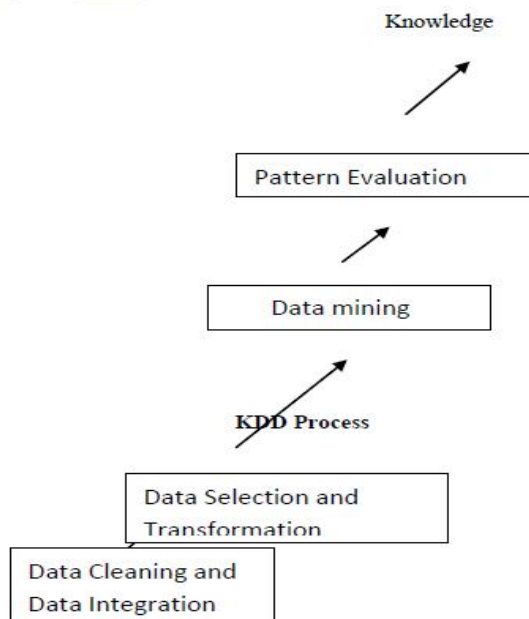
Fig.1 -KDD Process

The various techniques used in Data Mining are:
A. Association Analysis

Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data. Association analysis is widely used for market basket or transaction The Basic format of Association Rule is as follows :

Let A={l1,l2…lm} be a set of items .Let T be the transaction database be as set of transactions where each transaction t is a set of items. Thus, t is a subset of A. There are two basic measurements for each rule, support and confidence. A transaction t is said to support an item li is present in t. t is said to support a subset of items **X is Subset of A ,** if t supports each item l in X. An itemset **X is Subset of A** has a support s in T denoted by s(X)t, if s% of transactions in T support X.

B. Classification and Prediction

Classification is the processing of finding a set of models (or functions) which describe and distinguish data classes or concepts, for the purposes of being able to use the model to predict the class of objects whose class label is unknown.
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. However, in

many applications, one may like to predict some missing or unavailable data values rather than class labels. This is usually the case when the predicted values are numerical data, and is often specifically referred to as prediction. IF-THEN rules are specified as IF condition
THEN Conclusion
e.g. IF age=youth and student=yes then
joins computer=yes

C. Clustering Analysis

Unlike classification and predication, which analyze class labeled data objects, clustering analyzes data objects without consulting a known class label. In general, the class labels are not present in the training data simply because they are not known to begin with. Clustering can be used to generate such labels. The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. That is, 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. Each cluster that is formed can be viewed as a class of objects, from which rules can be derived. Clustering in distance education can help the organizations to group students into classes of similar and dissimilar.

D. Outlier Analysis

A database may contain data objects that do not comply with the general behavior of the data and are called outliers. The analysis of these outliers may help in fraud detection and predictingabnormal values.
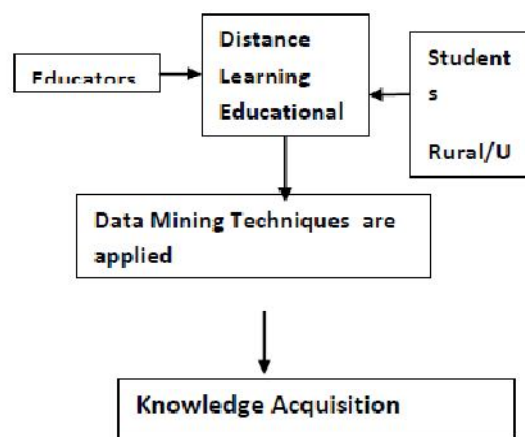

Fig.2 - Data Mining in Distance

Learning System
Data Cleaning and Data Integration
Cleaning and Data Integration Data Selection and Transformation Data mining Pattern Evaluation.

The 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 furtherhelps the educators and students to make decisions.

III. METHODOLOGY

Presently, distance educational institutes is to maintain a high quality educational programme which will improve the student's learning process they must also explore the courses that are in available for employability and optimize the use of resources. A typical student(rural/urban) at the university level completes a number of courses prior to graduation, organization of syllabi is influenced by many factors such as socio-economic constraints. Methodology follows as

- Identify the possible related courses and age wise preferences
- ☐Determine the strength of their relationships and Determine strongly related subjects.

Step 1 - Association rule mining is used to identify two course and age combinations
Step 2 - Pearson Correlation Coefficient was applied to determine the strength of the relationships of course and age combinations identified in the first step.

sno	Name	Break (in years)	Purpose of Drop-out	Course Prefer	Age
1	Rama	3	Financial	B.Com	21
2	Johny	10	Socio-economic	B.A	32
3	Teena	3	Marriage	PGDCA	22
4	Rita	2	Health	BCA	20
5	Maria	10	Employed	M.Sc	33
6	Alice	3	Back-logs	B.Sc	22
7	Bhanu	4	Marriage	MCA	26
8	Pavani	5	Children	M.A	30

TABLE -1. The Courses Chosen by students

Association Rules that can be derived from Table 1 are of the form:

$$age(X, "25..40") \wedge gap(X, "3..10")$$

$$\longrightarrow prefers(X, "PG Course")$$

[Support = 2% ,Confidence= 75 %]

Predicting The Registration of Students in an Educational Programme through Regression Analysis Now a day's distance educational organization are getting strong and needs in depth and enough knowledge to predict the registrations assessment on the latest courses framed , evaluation, planning which benefits both the students and the organization. Data Mining helps to identify the useful patterns to predict the performance with great accuracy which helps the universities to allocate resources. Regression analysis is a statistical tool for the investigation of relationships between variables. In research we are interested in ascertaining the causal effect of one variable upon another.

IV. CONCLUSION

In the present study, we have discussed the various data mining techniques which can support education system via generating strategic.

REFERENCES

[1] Han Jiawei, Micheline Kamber, *Data Mining: Concepts and Techniques,* Morgan Kaufmann Publishers,2000
[2] Sun Hongjie, ―Research on Student Learning Result System based on Data Mining, IJCSNS International Journal of Computer Science and Network Security, Vol.10, No. 4, April 2010
[3] Hamalainen, W., Suhonen, J., Sutinen, E. and Toivonen, H.: Data mining in personalizing distance education courses. World Conferenceon Open Learning and Distance Education, Hong Kong (pp. 1-11). (2004).
[4] Tarafdar, M .K. and Alam, S (2001)
Globalization of education through technology: some ethical and moral issues. Journal of Bangladesh Open University, vol.11 no.1

About The Author's

B Deepthi Chary . Completed M.Tech from JNTU in a Specialization Computer Science Engineering. Presently Working as Assistant Professor in MRIET, JNTU, and Hyderabad. Area of Specialization is Data Mining, Java and Network security

Y SUCHARITHA . Completed M.Tech from JNTU in a Specialization Computer Science Engineering. Presently Working as Assistant Professor in MRIET, JNTU, and Hyderabad. Area of Specialization is Data Mining, Complier design and Network security

P CHANDRASHAKER REDDY. Completed M.Tech from OU in a Specialization Computer Science Engineering. Presently Working as Assistant Professor in MREC, JNTU, and Hyderabad. Area of Specialization is Data Mining, Software engineering and Cloud computing