

Data Mining: A Path for Effective Counseling and Course Selection

Dharmendra Pal

*Associate Prof., Department of Information Technology
Dronacharya College of Engineering, Gurgaon-123506, Haryana, India*

Bhanu Pratap Saini

*Assistant Professor, Department of Computer Science
Shyama Pradas Mukherji College for Women, (University of Delhi), New Delhi-110026, India*

Abstract- Data mining technologies are used to discover the hidden information from the varying data sets. Data mining methodologies are broadly used in various real life business applications like Banking, Insurance, Medicine, Fraud detection, Product retailing, Security systems etc. data mining methods are used in business process optimization and control. Data mining involves the use of data analysis tools to discover previously unknown, valid patterns and associations in large data sets. The synchronization between data warehousing and data mining allows the information seekers to improve the quality and effectiveness of their decisions.

The biggest challenge in professional education is predicting the paths of students and alumni. Institutions would like to know that which students will enroll in particular course programs, and which students will need assistance in order to graduate. One way to effectively address these student and alumni challenges is through the analysis and presentation of data, or data mining. Data mining enables organizations to use their current reporting capabilities to uncover and understand hidden patterns in vast databases. The patterns are then built into data mining models and used to predict individual behavior with high accuracy.

Keywords- Data Mining, knowledge management, audio, genetic algorithms, Pattern, Data mining models.

I. INTRODUCTION

Data mining involves some complex data analysis tools to discover the previously unknown valid patterns and relationships in a large data sets. The tools can be mathematical algorithms, statistical methods, learning algorithms etc. Data mining is a tool of business intelligence for knowledge discovery. The predictive power of data mining[1] comes from concepts of machine learning, pattern recognition, and statistics and it automatically extracts concepts to determine interrelations and patterns of interest from large databases. These techniques and methods that are used in industries can be well connected to academia.

Data mining is a powerful tool for academic institution. Every Academic institute carry three duties that are data mining intensive: scientific research that relates to the creation of knowledge, teaching that concerns with the transmission of knowledge, and institutional research that pertains to the use of knowledge for decision making. They expend a lot of budget towards information technology infrastructure in order to compete with other institutions. So it becomes even more important for the institutions to achieve success with benchmarking quality. Often the institutions come across several patterns in evaluations, courses, students' counseling and admissions. These patterns if extracted by using data mining techniques will enhance data sharing analyze diversified student relationship management[2] & predict student performances and success of programs.

To take advantage of the results, a system needs to be in place (Tsantis and Castellani, 2001) for transforming new knowledge into successful models for teaching and learning to develop and improve student relationship management. Luan (2001) emphasizes on knowledge management for higher education and emphasizes the role of data mining in research, teaching and institutional research.

II. REQUIREMENT OF DATA MINING

Extracting the useful information from the varying data sets is very complicated process. There is lot of requirements in data mining [3].

2.1 Handle different type of data: Many databases have different types of complex data like audio data, video data, spatial data, and hyperlink data. Data mining technique must be able to perform on various types of data structures. It is impossible to expect a data mining technique to handle all types of data. In general a specific data mining technique is used to mining the knowledge from a specific type of data.

2.2 Information from different types of sources: In today’s world there are different data sources available like internet, intranet data warehouse where data is stored in different format. The data mining algorithms should be capable to handle the data from different sources. Different levels of analytical tools are available for this purpose.

2.3 Artificial neural networks: Neural network refers to the human brain by learning from a training dataset and applying the learning to generalize patterns for classification and prediction. These algorithms are effective when the data is shapeless and lacks any pattern. The basic unit of an artificial neural network is modeled after the neurons in the brain. This unit is known as a node and is one of two main structures of the neural network model. The other structure is a link that corresponds to the connection between neurons in the brain.

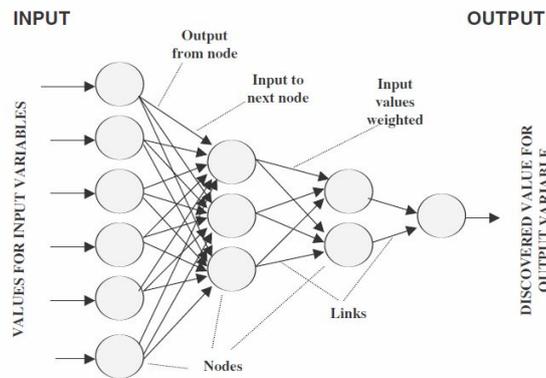


Fig1: Neural Network

2.4 Genetic algorithms: Genetic algorithm technique is based on the neural neural network technology. Over generations, the process propagates the genetic material in the fittest individual from one generation to the next. Genetic algorithm apply the same principles to the data mining. This technique uses highly iterative process of selection, cross-over, and mutation operators to evolve successive generations of models. At each iteration every model competes with everyone other by inheriting traits from the previous ones until only the most predictive model survives.

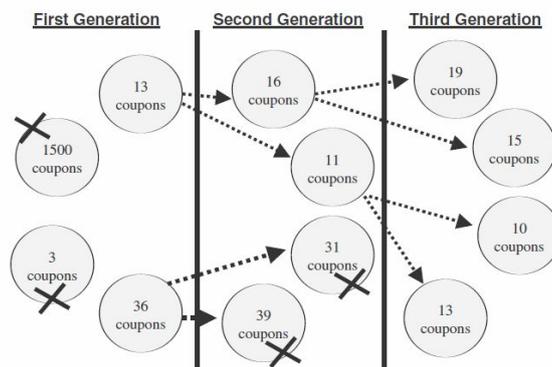


Fig2: Genetic Algorithm

2.5 Decision trees: A decision tree is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and leaf nodes represent classes or class distributions. The topmost node in a tree is the root node. By following the tree, you can decipher the rules and understand why a

record is classify in the certain way. These decisions generate rules for the classification of a dataset. The Figure shows a decision tree representing the profiles of men and women buying the notebook.

Decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

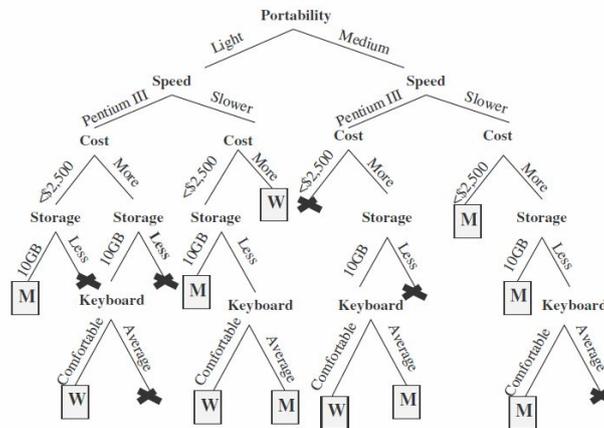


Fig3: Decision Tree

2.6 Link Analysis: Link analysis is used to find the patterns from the relationships and discovers the knowledge. Ex. Airlines link cities together. Telephone calls connect people and establish relationships. By analyzing the relationships between two phone numbers established by the calls along with other stipulations, the desired information can be discovered.

III. KNOWLEDGE MANAGEMENT

Knowledge is the process of moving the data to information. The components of knowledge management [4] are Explicit (Measurable) and Implicit (Qualitative). Explicit knowledge is exists mainly in the form of database and always available to the user. Implicit knowledge is very hard to quantify. It can be personalities, feelings or aptitude etc. CRM (Customer Relationship management) uses both types of knowledge work together for the effective business solutions. The following diagram illustrates the modern component of knowledge Management.

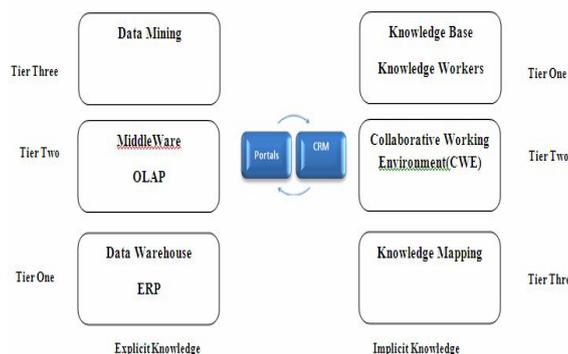


Fig4: Knowledge Management Framework

Explicit data mining reflects the highest level of knowledge management which requires data domain skill, data querying and presentation skill and artificial intelligence/machine learning skill .Data mining resides at the Tier

Three and depends on the lower tiers. The following diagram reflects the Explicit Knowledge framework which shows the relationships between the three tiers and the software programs required for each level.

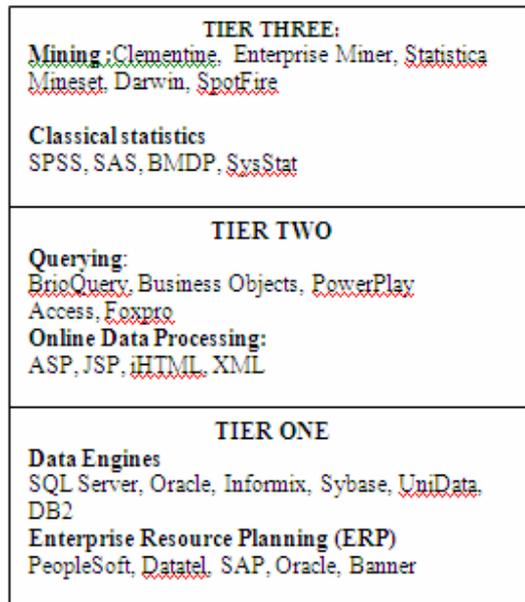


Fig 5: Data mining for Explicit Knowledge

IV. PROFESSIONAL EDUCATION STUDENT COUNSELLING & SELECTION PROCESS

The student gets admitted in to any higher education institutes based on their performance in various competitive exams like IIT, AIEEE, MBBS or IIM's and various other universities. The selection is based on the highest score obtained by student in the entrance exams. Among the entire program the admission process includes the students' demographic and academic data is taken to validate the admission and selection process and effectiveness of the counseling program.

The counseling process starts when student is selected by the admission process and get registered. The major problem lies in solving the problem of a particular stream may be opted by majority of the student that increases the load on the faculty and student as well as institution. The quality of professionals finally produced by an institute depends on the quality of 3Cs(Counseling for choice of correct program, Counseling for choice of correct Subject and Counseling for choice of correct job).

Data mining approach to academic related data [5] is easily adaptable to different types of courses, different population sizes, and allows for different features to be analyzed. We know that every student has unique feature in terms of capabilities and preferences. So in order to get the correct choice the process must be backed with some kind of data support and data analysis. Ex. whether a student with under-graduation in commerce should take up marketing or finance as a stream of management course is not a linear decision based on some obvious relation between commerce and finance. The decision is actually multi-factor, multi-criteria. These processes can be improved by understanding the integrated view of academic data and applying appropriate data mining techniques to it.

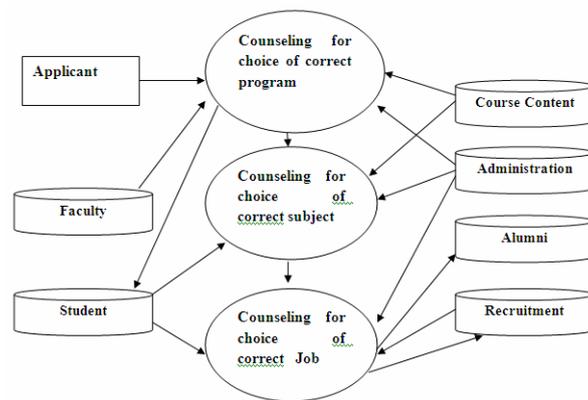


Fig6: Counseling Process

I am interested in analyzing the data for pattern like student interaction, admission process, student counselling mechanism, list of courses, subjects offered, placement pattern etc.

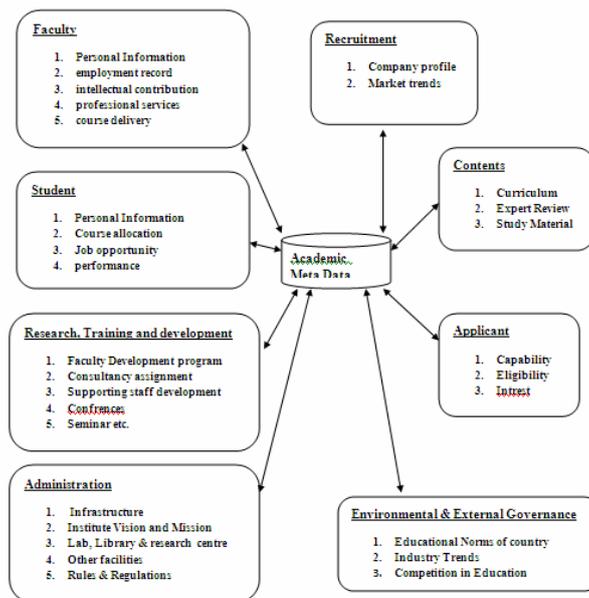


Fig7: Data resource View

V. DATA MINING CHALLENGES

There are several issues related to the implementation of Data mining. This can include data quality, interoperability, misuse of data, and privacy. There are other factors also that influence the project's outcome.

5.1 Quality Data: Data quality defines the accuracy of the data. Data quality can also be affected by the structure of the data. The presence of duplicate records, lack of data standards, timeliness of updates and human error can impact the effectiveness of the data mining techniques. To improve the data quality it is necessary to clean the data. This requires the deletion of duplicate records, normalizing the values in the database (ex. No is represented by 0

throughout the database), accounting for missing data points, identifying anomalous data points (Ex. Age as 150 years) and standardizing data format.

5.2 Interoperability: Interoperability refers to the ability of a computer system and/or data to work with other systems or data using common standards or processes. Interoperability of databases and software is important to enable the search and analysis of multiple databases simultaneously, and to ensure the compatibility of data mining activities of different agencies. Data mining projects that are trying to take advantage of existing legacy databases or that are initiating first-time collaborative efforts with other agencies or levels of government (e.g., police departments in different states) may experience interoperability problems. Similarly, agencies may require the creation of new databases and information sharing efforts, and then they will need to address interoperability issues during their planning stages to better ensure the effectiveness of their data mining projects.

5.3 Misuse of Data: This refers to the use of data for purposes other than that for which the data was originally collected. This can occur regardless of whether the data was provided voluntarily by the individual or was collected through other means.

One of the primary reasons for misleading results is an inaccurate data. All data collection efforts suffer accuracy concerns to some degree. Ensuring the accuracy of information can require costly protocols that may not be cost effective if the data is not of inherently high economic value. In well-managed data mining projects, the original data collecting organization is likely to be aware of the data's limitations and account for these limitations accordingly. However, such awareness may not be communicated or heeded when data is used for other purposes.

For example, the accuracy of information collected through a credit card may suffer for a variety of reasons, including the lack of identity authentication when a card is issued. For the purposes of marketing to consumers, the impact of these inaccuracies is negligible to the individual. If a government agency were to use that information to target individuals based on credit history associated with particular region, then the result will be based on inaccurate information could be, at the least, a waste of resources by the government agency, and an unpleasant experience for the misidentified individual.

5.4 Security and Data Privacy: The potential of data mining applications to can be extended beyond their original purpose. So the privacy concerns with both actual project and proposed projects. The degree to which government agencies should use and mix commercial data with government data, whether data sources are being used for purposes other than those for which they were originally designed, and the possible application of the Security Act.

VI. CONCLUSION

Data mining is very powerful tool for decision making and learning process through the information discovery. By combining data mining with student demographic and other information enables the educational institutes to improve the understanding of the various categories of student. Data mining combines the best practices of the information technology and academic processes through efficient models, analyzes the diversified student relationship management, assesses the alumni affairs, etc. Survey says that Data mining will be a critical factor for the business over the coming year [6]. Data mining applications like predicting student performances and success of programs, automated alerts and what if analysis using data mining will be extremely helpful in academic learning environment. It will help the educational institutes to meet the requirements if various student groups.

REFERENCES

- [1] Edelstein, H. "Data mining: exploring the hidden trends in your data", DB2 Online Magazine. (1997).
- [2] Luan, J. Data Mining Applications in Higher Education, Jossey-Bass, San Francisco, CA (2001).
- [3] Lu, H. , Setiono, R. and Liu, H. "Data mining: an overview from database perspective" IEEE transaction on data and knowledge engineering. Vol. 8. No. 6. Pp. 957-61. (1996)
- [4] Fayyad, U., Djorgovski, S. G. and Wair, N. "Automating the analysis and cataloging of sky surveys, advances in knowledge discovery and data mining ", MIT press.

- [5] Minaei-Bidgoli, B. and Punch, W.F. ,“Enhancing online learning performance: an application of data mining methods”, The 7th IASTED International Conference on Computers and Advanced Technology in Education (CATE 2004), Kauai, Hawaii, USA, August, pp. 173-8. (2004).
- [6] Baker, S. and Baker , K. “Mine Over Matter”, Journal of Business Strategy, Vol 19 No 4. Pp. 22-27.