

# A Review on Classification and Prediction Based Data Mining to Predict Slow Learners in Senior Secondary Schools

Sohajbir Singh Ubha

Rayat Bahra University, Mohali

Punjab, India

Gaganpreet Kaur Bhalla

Rayat Bahra University, Mohali

Punjab, India

---

**Abstract:** One of the biggest challenges that higher education faces today is predicting the paths of students. Institutions would like to know, for example, which students will enrol in which course, and which students will need assistance in particular subject. The new interesting subject that is being offered by institution to interact more students. How management get better information about student, their result, about the success of new offered courses, the answer is data mining .it helps to institution to take decision .data mining is the better tool to predict the result of the student.

In this paper, I will discuss about data mining their different phases, advantages of data mining in the field of academics, some results and tools used for data mining and their applications.

**Keywords:** KDD (Knowledge Discovery in Database), J48 Algorithm, Naïve Bayesian Classifier, WEKA data mining tool.

---

## 1. INTRODUCTION

We Data Mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help universities or institutions to focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing institution 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. Data mining tools can answer institution questions that traditionally were too time consuming to resolve [5]. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. Data mining is a powerful tool for academic intervention. Through data mining, a university could, for example, predict with 85 percent accuracy which students will or will not graduate. The university could use this information to concentrate academic assistance on those students most at risk.

In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: classification, categorization, estimation, clustering and visualization [1]. Classification identifies associations and clusters, and separates subjects under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay." Estimation includes predictive functions or likelihood and deals with continuous outcome variables, such as GPA and salary level. Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. Visualization is used primarily to depict three-dimensional geographic locations of mathematical coordinates [2]. Higher education institutions

can use classification, for example, for a comprehensive analysis of student characteristics, or use estimation to predict the likelihood of a variety of outcomes, such as transferability, persistence, retention, and course success.

### Data mining tools and algorithms

- Machine learning
- Computer science, heuristics and
- Induction algorithms
- Artificial intelligence
- Emulating human intelligence
- Neural networks
- Biological models and Engineering

### Phases of data mining

Data mining is an iterative process that typically involves the following phases:

- Problem definition
- Data exploration
- Data preparation
- Modeling
- Evaluation

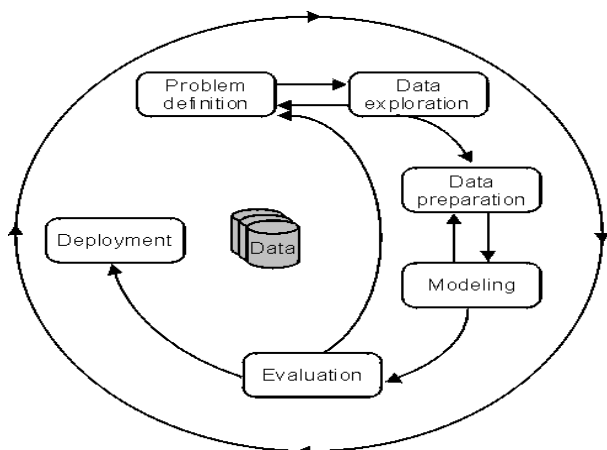


Figure 1 Process of Data Mining

A data mining project starts with the understanding of the problem. Data mining experts, business experts, and domain experts work closely together to define the project objectives and the requirements from a business perspective [4].

In my project our domain is academic data like a student records, result of colleges of different years, strength of students per year and per department and the experts of that domain are HOD's of departments and principals of the colleges.

In the data exploration phase, traditional data analysis tools, for example, statistics, are used to explore the data. In the data preparation phase, data is tweaked multiple times in no prescribed order. Preparing the data for the modeling tool by selecting tables, records, and attributes, are typical tasks in this phase. The meaning of the data is not changed [3].

We select and apply various mining functions because we can use different mining functions for the same type of data mining problem. Some of the mining functions require specific data types.

In the modeling phase, a frequent exchange with the domain experts from the data preparation phase is required.

Evaluate the model. If the model does not satisfy their expectations, they go back to the modeling phase and rebuild the model by changing its parameters until optimal values are achieved. When we are finally satisfied with the model, we deployed it.

### Tools of Data Collection & Analysis

Various tools are needed for that project some for analyzing data, some for designing, implementation and some developing software tool these are:

- Excel
- Ms access
- J48 algorithm
- Naïve Bayesian Classifier

- WEKA data mining tool
- Tangara data mining tool
- Rapid miner

### Advantage of Data Mining In Academics

Data mining tells us following things like:

- Tells us about the weak students.
- Tells us which students are taking more credit hours.
- Subjects which are more interesting to students.
- Type of courses we can offer to attract more students.
- Tells various ways to help weak students.
- Helps in improving the result of schools.
- Helps in predicting the result of students.

## 2. CONCLUSIONS

The current education system does not involve any prediction about fail or pass percentage based on the performance. The system doesn't deal with dropouts. There is no efficient method to caution the student about the student about the deficiency in attendance. It doesn't identify the weak student and inform the teacher. Another common problem in larger colleges and universities, some students may feel lost in the crowd. Whether they're struggling to find help with coursework, or having difficulty choosing (or getting into) the courses they need, many students are daunted by the task of working through the collegiate bureaucracy. Since the proposed model identifies the weak students, the teachers can provide academic help for them. It also helps the teacher to act before a student drops or plan for recourse allocation with confidence gained from knowing how many students are likely to pass or fail. Proposed system also shows data graphically according to the need or organization which help them to take important decisions. For future work we also use clustering, with the help of clustering we can see the domain and interest of students in particular field [1].

## 3. REFERENCES

- [1]. Hideko Kitahama, "data mining through cluster analysis evaluation on internationalization of universities in japan".
- [2]. Bruce I. Golden r. H. Smith School of Business University of Maryland, College park, md 20742 "an example of visualization in data mining"
- [3]. Jing Luan, PHD chief planning and research officer, Cabrillo College founder, knowledge discovery laboratories "Data Mining Applications in Higher Education".
- [4]. Thulasi, kumarthulasi.kumar@uni.edu, university of northern iowa "theoretical basis for data mining approach to higher education research".

- [5]. N.V.Anand Kumar Research Scholar, Department of Computer Science and Engineering, Anna University, Chennai “Improving Academic Performance of Students by Applying Data Mining Technique”.
- [6]. Han, J. W., Kamber, M., 2006. Data Mining: Concepts and Techniques, 2nd Edition, The Morgan Kaufmann Series in Data Management Systems, Gray, J. Series Editor, Morgan Kaufmann Publishers.
- [7]. Luan, J., 2002. Data mining and knowledge management in higher education – potential applications. In Proceedings of AIR Forum, Toronto, Canada.
- [8]. D. A. Alhammadi and M. S. Aksoy, "Data Mining in Education – An Experimental Study," International Journal of Computer Applications, vol. 62, no. 15, pp. 31-34, 2013. [8]. B.K. Bharadwaj and S. Pal. “Data Mining: A prediction for performance improvement using classification”, International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.
- [9]. Alaa el-Halees (2009) Mining Students Data to Analyze e-learning Behavior: A Case Study.
- [10]. U. K. Pandey, and S. Pal, “Data Mining: A prediction of performer or underperformer using classification”, (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN: 0975-9646, 2011.
- [11]. U. Fayyad, Piatetsky, G. Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0–262 56097–6, 1996.
- [12]. Berry, M.J.A., & Linoff, G.S. (1997). Data Mining Techniques for Marketing, Sale, and Customer Support, New York: John Wiley & Sons, Inc.
- [13]. Chen, S.Y. (1999). Data Mining in Acquiring Association Knowledge, Between Diseases and Medicine Treatments. Unpublished Master’s Thesis, National Sun Yat-Sen University, Kaohsiung, Taiwan R.O.C.
- [14]. Codd, E.F. (1993). Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate. E.F. Codd and Associates.
- [15]. Dai, C.Y. & Sung, J.K. (2005). Data Warehouse of TVC Course in Taiwan, ED-MEDIA 2005, Montreal, Canada, June 27-July 2,2005, AACE.
- [16]. Executive Yuan (2002). Promotion Program for Strengthening Digital Content Industry, Executive Yuan of R.O.C.
- [17]. Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery: An Overviews. Advances in Knowledge Discovery and Data Mining, Menlo Park, CA: AAAI Press.
- [18]. Foley, J. (1996). Squeezing more value from data, Information Week, (December 9, 1996): 44.
- [19]. Henderson, D. (1996). Tackling an OLAP development project, Computing Canada, 22(19), 24-25.
- [20]. Liao, I.Y. (2002). The study on application of data mining for pharmaceutical market in Taiwan, Unpublished Master’s Thesis, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C.
- [21]. Greenfeld, N. (1996). Data Mining, UNIX Review, 14(5), 9-14.
- [22]. Peacock, P R(1998). Data Mining in Marketing: Part 1, Marketing Management, 6(4), 8-18.
- [23]. Rob, P & Coronel, C. (2004). Database Systems Design, Implementation & Management, Boston, MA: Course Technology, Thomson Learning, Inc.