

Data Mining Aspects for Student Performance and Analysis in Web-Based Educational System LON-CAPA

Ritika^{1*} Dr. Kalpana Midha²

¹ Research Scholar, OPJS University, Churu, Rajasthan

² Associate Professor, Department of Computer Science, OPJS University, Churu, Rajasthan

Abstract – Data mining, analytics & machine learning are used on EDM information to extract information from educational environments. It is increasingly in demand and draws further interest because of the rise in educational data in web based learning and even the development in traditional education. This indicates that the potential of data mining tools to provide novel solutions for decision-makers to solve problems exists in particular areas. The data analyzed and the knowledge collected from the educational field using DM techniques was referred to as educational data mining (EDM). This research offers a tool to classify student features in order to predict test performance based on attributes derived from historical data within a web-based educational system. We will propose a specific version of the laws of contrast as well as a method for the detection of such trends. This approach is applicable to the LON-CAPA method.

Keywords – Data Mining EDM LON-CAPA, Intelligent Tutoring System

-----X-----

INTRODUCTION

Data mining, analytics & machine learning are used on EDM information to extract information from educational environments. It is increasingly in demand and draws further interest because of the rise in educational data in web based learning and even the development in traditional education. Several educational institutions are seeking to create on-line computing and research proximity. A variety of electronic communication platforms with diverse capacities and strategies have been set up to offer online schooling in a scholarly environment. Actually, Michigan State University (MSU) has pioneered a number of these programs to establish a basis for electronic instruction. The research presented here was carried out on a piece of the most current online education program built at MSU, the Learning Online Network with Computer-Assisted Personalized Approach (LON-CAPA). This framework outperforms the board programs in three main aspects of learning assessment. The first is the capacity to individualize problems, including algorithmic computational operations and things that are contextual and implemented. The second is that it helps educators to collaborate together in the development and exchange of content in a simple and productive way, both inside and through foundations. That is more, the third is its special objective power: that is, its power to turn a particular educational commodity,

e.g. a numerical or theoretical schoolwork query, into a structure that is appropriate for specific occupations.

The main goal of data mining is to discover a hidden connection between objectives in a data collection. The Web-based education techniques permit to teachers concentrate undergraduates study and which programs are better for learning. Once large measurements of under-study profile data are obtained by LON-CAPA, data mining and disclosure strategies can be used to discover fascinating relationships among under-study characteristics, tests, and under-study solution systems. The education sector has adopted this technology and developed a web-based education system.

DATA MINING

History of LON-CAPA

In the fall of 1992, CAPA (Computer-Assisted Customized Approach) was directed in a limited material science class of 92 undergraduate studies. CAPA provides under-studies with customized sets of issues, tests and tests. Understudies are supported with momentary feedback and observations through the Web and through fix blunders without penalty before the due date of the mission. The program tracks the engagement and

success of the undergraduates, and the documents are available electronically to both the instructor and the undergraduate. CAPA is a demonstration apparatus, not an educational plan, and all things considered are not direct course configuration, substance or objectives. Instead, it empowers workers to extend their courses with relevant, individualized events.

The Lecture Online project started to represent material science course content on the internet in the fall of 1997 with 770 understudies. With only the web-program as an interface, Lecture Online empowers teachers to constantly put together the materials collected from all over the Internet and to make various kinds of individualized online school work. Evaluation, correspondence, group work and enrolment are additionally handled by the system.

In 1999, CAPA and Lecture Online joined forces to set up the CAPA (LON-CAPA) Learning Online Network, which provides a superset of CAPA and Lecture Online functionalities. LON-CAPA was first beta-tested at MSU in the spring of 2001.

Currently, LON-CAPA is used in more than 160 other institutions, varying from school to university level courses. Controls include space science, science, business, science, structural design, computer science, family & youth, topography, human nutrition & sustenance, human medicine, mathematics, clinical innovation, material science, and brain research.

LON-CAPA, System Overview

LON-CAPA is a common teaching administration program that gives under-considers customized collections of issues, appraisals and tests. Personalized (or individualized) schoolwork ensures that any understudy sees a rather remarkable computer-created problem. LON-CAPA offers understudies and instructors with appropriate insight into analytical comprehension and approach precision. In addition, it enables employees to expand their courses through individualized, significant activities and to create and share specific online assets. LON-CAPA implies that this value is put on a homogeneously passed on stage for the creation, sharing & passing on obviously content by accentuation on cross-institutional cooperation & the security of administrators' development rights.

LON-CAPA Topology

LON-CAPA is designed as a network of constantly connected servers that are scattered geographically. Both devices on the network are connected by careful two-way TCP / IP connectivity. The server consists of two client groups: library servers, or control servers. A library server can act as a home server that holds all of the user's personal records and is responsible for initial user authentication when

accessing a session on any computer on the network. It also houses the writers' development region and the official copy of any document written by the poet. The Access Server is a student conference processing system. Library Administrations can be used as a substitute for meeting management when all network entrance servers are overpowered.

Through LON-CAPA client is a member of a solitary field. Areas can be distinguished by disciplinary or administrative boundaries, such as MSU, FSU, OHIOU and the identity of distribution agency. These territories may utilized to restrict the advancement of specific client subtleties transversely the network, assign privileges & implement sovereignty plans. As a result, the understudy of subtleties and course data are dispersed over various vaults. Every gadget consumer has a library account, which is his or her home machine. It houses a complete set of all their documents.

LON-CAPA actually has a huge rise in demand for Redhat-Linux Intel-good hardware. LON-CAPA actually has a huge rise in demand for Redhat-Linux Intel-perfect hardware. The present MSU configuration comprises of a few connection & library servers. Both entrance servers are fixed as front-end devices on a mutual IP conspirator and are subject to the "application conference" understudy. The new LON-CAPA execution utilizes mod perl within the Apache web server programming.

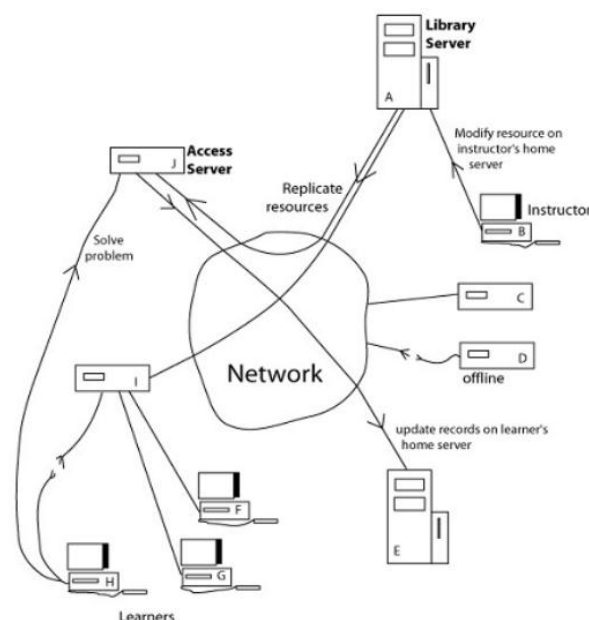


Figure 1.2 A schema of distributed data in LON-CAPA

Learning Outcomes

The greatest test of progress for our program is that it truly appears to be an successful mechanism for rising educational opportunities for students. A

number of empirical studies, conducted predominantly in the context of undergraduate physics courses, suggests that LON-CAPA may have a substantial impact on student performance. One study (Kashy, 2001) followed an advanced physics course focused on calculus from years before the launch of the software until late in implementation. For the years before LON-CAPA was implemented, the actual grade range displayed a typical bell form of grade 2.5, with very few students earning grades 3.5 or 4.0. Since transferring to LON-CAPA, the percentage of students receiving a higher rating than 3.0 improved significantly. In fact, professional evaluators noted that the assessments included in the post-deployment course were more difficult than those utilized in previous years, and that the significant improvement in educational results cannot be due to the reduction of class levels. Two separate investigations indicate that LON-CAPA can expand women's participation and achievement in science. One study (Thoennessen, 1996) in view of a year-long course in material science for non-science majors, in which the curriculum was used in the following semester in a special way. Total grades from the following semester showed women are expected to benefit from the program. The ensuing analysis found that ladies, who started the course slightly less prepared than men, increased their yield over the semester in comparison with men until there were no sex differences at the end of the year study. These early findings are being explored in greater detail as of now. Most understudies, usually 80 percent, notice that they are told by LON-CAPA, and welcome the material of the course. It is surprising, as time spent on homework and other course requirements by understudies has gone up by a factor of two by far.

Unanticipated Challenges

LON-CAPA has an proven track record of the use of student education technologies. A recent trend has been that learners are questioning educators by using technology. Although the LON-CAPA configuration feature still restricts the red copying of responses rather than any other software, students have created collaborative websites (funded by banner ads and donations) to network with each other in an attempt to beat the system. In an effort to re-engineer individualization processes, entire Excel spreadsheets are being created—one problem at a time, containing 52,359 Homework Platform Comments. In addition, these programs have encouraged educators to create even more nuanced tasks, including accidental labeling on individual graphics. The final word on this "online war" has not yet been released, but a study in which students were asked at the end of the course how much they used the hacking platform found the final score to be highly adversely correlated with platform use.

Sustainability and Positioning

LON-CAPA is one of a kind that is developed under the open-source GNU General Public License between teaching and learning content management systems. We do not believe that the designers community at Michigan State University alone will have the option at continuing the LON-CAPA code base (currently about 60,000 lines of code) without the involvement of a more comprehensive network of improvements. Unlike prohibitive business programming, the benevolent open-source network is anything but hard to concentrate on researching and adapting to software growth (as delineated by the Linux working framework). We are looking at whether this concept is convincing for a system that operates on the scholastic network, where software developers themselves are not the fundamental accomplices when in question. However, as of today, the LONCAPA system has a code base that includes commitments from four colleges and joins countless accessible open-source devices and libraries. LON-CAPA has never been expected to compete in the Executive Systems Business Course, and offers a naturally scholastic approach to online education and learning. The allocation of benefits relies on the buddy summary model (through the selection of benefits by a partner and the giving of credit to the maker) and will formalize this concept with an specifically peer-evaluated subset of the benefit pool before long. It empowers instructors to conduct instructional tests on a colossal scale and to watch the instructional impact of directed interventions by prudence of its ability to be produced and balanced, even as its wide-ranging account of understudy gets in touch with them. At the annual Frontiers of Education Conference (1997, 1998, 2000), the Ben Dasher Award (1998), the Wickenden Award (1999), and the Computer World Honors Award (2003), IEEE and ASEE honored this effort with three Best Paper Awards. This job has consistently sponsored ordinary gatherings and meetings for a number of members to grow. At the 2002 LON-CAPA Client Meeting, organized by Florida State University, 56 staff from 22 organizations typically participated as researchers. 72 agents talked to 28 groups at the Truckee Meadows Community College in Reno, Nevada, during the 2003 meeting. LON-CAPA offers them an opportunity to speak about educational problems, for which they would not usually have a platform. The extension of the LON-CAPA Server Network is critical both for the timely deployment and for the long-distance manageability of the undertaking. Started as a basic scholarly endeavor, the program would, in the long run, show the cash-related support without the grant funding. This is a challenge to render a transition without compromising one of the kind attributes of the program and without turning a scholarly office of

work workers into an entirely cash suspense oriented assistance material.

Intelligent Tutoring System

While the word intellectual tutoring program is commonly used in study literature, given all, more work focuses on elective words. This variation of terminology may contribute to a failure to comply with the relevant study obligations (Steenbergen-Hu and Cooper, 2014). In addition, the lack of consensus on a binding together sense of ITS can often result in perusers becoming responsive to suggestions arising from this audit. It is therefore necessary to establish an unmistakable description for ITSs. Numerous attempts have been made to distinguish ITS from other computer-based instructional programs on the grounds of its capacity to deliver modified instructions. In 1982, Sleeman and Brown characterized ITSs as "a versatile system that uses intelligent technologies to tailor learning to unique qualities, such as information on matter, state of mind and feeling" (p. 2). Shute and Psotka (1996) have claimed that for the ITS, "the most important aspect is real-time conceptual analysis (or understudy modeling) [and] scalable remediation" (p. 14). In 1999, Self claimed that "ITSs are computer-based learning methods that aim to respond to the needs of students and are thus the key structures that try to 'protect' students in that way" (p.350). More recently, Conati (2009, p.2) indicated that "ITS is an interdisciplinary area that examines ways to build instructional programs and offer instruction relevant to the needs of particular students, with the same amount of successful educators." In addition, Pacella (2014) also stressed that, despite having mastery in the matter, the ITS must have the option of storing under-information, including prior information and progress, regardless of directing important assessment for each progression and introducing customized learning materials for under-

Despite the fact that ITSs differ in UI highlights, subject spaces and learning factors demonstrated, some basic structural segments serve as the center of usefulness to demonstrate under-studies by adjusting to changes in information. In the primary ITS, Teacher, Carbonell (1970) developed three core modules to support customization: the master module, the student module and the guidance module. One of the three segments was joined together to create limited combined behavior training transfers. Most early ITSs adopted Carbonell's three-segment concept (e.g. Polson and Richardson, 1988). Later, Dede (1986) expanded this model by indicating the UI as the fourth module in the ITS. While the UI is not necessarily another element, basically, like every ITS relies on it to specifically impart and interact with understudy and to collect understudy data to be included in other modules throughout the program, the identification of the UI as a stand-alone component of the ITS architecture has its importance throughout promoting a smooth learning experience. This also illustrates the growing

unpredictability of the client activities found in the further evolved ITSs. Within the conceptual ITS, the control engine produced a description of the space model and the understudy model to analyze the various learning experiences of the students. In addition, Sani and Aris (2014) have suggested comparable four-module Software architecture, which facilitates collaboration with under-studies through an interface framework and the collection of data used to modify and configure instructional assistance.

- Domain model
- Student model
- A tutor model
- A user Interface module

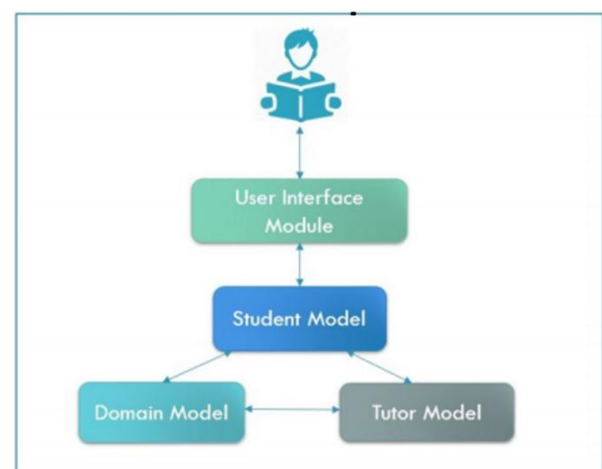


Figure 1.5: A Typical Architecture of an ITS

LITERATURE REVIEW

Minaei-Bidgolim et al. (2003) is amongst the key developers who defined students by utilizing GAs to determine their final score. Using relapse approaches, Kotsiantis and Pintelas awaited an understudy of imprints (pass and bomb classes). Planned scholarly achievement under-study (arranged into small, medium, and high-chance classes) utilizing distinctive DM approaches (decision trees and neural systems). This research provides a fruitful analysis for the evaluation and expectation of student success in higher learning environments utilizing DM advances.

Al-Radaideh et al. (2005) have developed a decision tree replica to forecast the absolute grade of undergraduate studies that have studied C++ at Yarmouk University, Jordan. Romero et al. have taken the lead on the updated DM methodologies in order to predict a final decision that will focus on the knowledge given by the e-learning system.

Khan et al. (2005) displayed a method to examine the presentation of the senior auxiliary understudies

dependent on the bunching plan. From the outcome, it was seen that young ladies with high socio-economic status had similarly higher scholastic accomplishment in science stream and young men with low socio-economic status had genuinely higher scholarly accomplishment when all is said in done.

Mazza and Milani (2005) presented a GISMO / CourseVis device that highlights procedures for the representation of information that can be used graphically to render multidimensional, complex under-studies following data collected by web-based education systems. Such techniques allow a significant volume of knowledge to be broken down by referring to details in certain visual presentation cases. Mostow used the Listen method to grasp their students and to be conscious of what's going on in the division classes.

Damez et al. (2005) used a fuzzy tree of preference for company modeling and thus separated the subject from the consumer studied. They use an operator to obtain experience in the psychological characteristics of a client's partnership and to classify clients as being checked or not.

Al-Radaideh, et al. (2006) functional a decision tree form to predict the final grade of undergraduate studies that inspected the C++ course at Yarmouk University, Jordan in 2005. Three distinctive grouping techniques in particular ID3, C4.5, and the NaïveBayes were utilized. The result of their outcomes showed that Decision Tree model would do well to prediction than other models.

Borstorff et al. (2007) The performance analysis showed that integrating more than one classifier has significantly increased the classification performance. A method has devised by predicting students' academic performances by using six different types of machine learning approaches for finding the efficiency of distance learning education, which was very much different from the regular education system. The performance analysis showed the performance features selected for the analysis and the demographic attributes involved in the process of better predicting capability were the major factors in machine based learning systems. In order to predict the student test score, a regression model is established in this document. This means that the best mixed-effect representations as aligned with the Bayesian network occur.

Bray et al. (2007) concentrated on private tutoring based performance analysis of the students in India compared to other countries. The study showed that the percentage of students preferred private education, which is comparatively high in India than other countries and it was also noticed that the performances of students was enhanced and the intensity was based on the socio-economic criteria.

Merceron and Yacef, (2007) used standard SQL questions for mining under-studied data taken from web-based tutoring and association guidelines, and emblematic data review. Their aim is to discover the botches that occur together on a daily basis. Becker introduced sequential trends that reveal which substance has given rise to the entrance to another substance, or how the apparatus and substance have become entangled in the learning process. Avouris et al. developed naturally created log documentation by posing the related details as an incremental resource and by supplying partner feedback and static records.

Rosi et al. (2009) Used a guided simulation, which is a sophisticated methodology for assessing the success of instructors in higher learning institutions, and suggested an appropriate algorithm and developed a program architecture suitable for forecasting the output of instructors, as well as advised the required action to be taken by school administrators in decision-making, taking into account the shortcomings of classical meth. The proposed system, once fully implemented, would support school administrators in decision-making; provide basis for instructors' performances improvement that will optimise students' academic outcomes and improve standard of education. Consequently, this will contribute to successful achievement of the goals.

Alaa Mustafa et al. (2009) Educational data mining was related to the mining process of educational data. Educational data mined for the purpose of evaluating the student's performance. This work focused on the process of analyzing the learning behavior of the students with EDM. In this a case study was reviewed. The data of the case study was collected from the course database. Then pre-processing techniques were applied to the data to make it more refine and usable. Then various data mining techniques were applied for the purpose of classification and clustering and detecting the outlier values. Then the behavior of the student was analyzed and described on the basis of above calculated parameters.

Sabourin et al. (2009) in this work author raised the concept of security in data mining. Here security referred to the security of the student related data from non-interesting organizations and un-authentic personnel. The student's data or academic data of the students was used by various business organizations for the purpose of surveys. Then it enhances the risk of security of the data. In this paper, author represents three criteria's on which the security was dependent. These main criteria were policy, corporate-socio responsibility and public interest or opinion. This paper provides the solution to the problem of security related to these fields.

Crist'obal Romero et al. (2010) Educational Data Mining (EDM) is an increasing interdisciplinary research area that handles the development of techniques to analyze data beginning in an educational setting. EDM utilizes computational methods to cope with dissecting educational data in order to understand educational inquiries. This paper studies the most applicable investigations did in this field to date. To begin with, it presents EDM and portrays the various gatherings of client, sorts of educational situations, and the data they give. It at that point proceeds to list the most average/basic undertakings in the educational condition that have been settled through data-mining strategies, lastly, the absolute most encouraging future lines of research are discussed.

CONCLUSION

This review uncovers a number of future lines of research in various measurements. EDM has become out of the current disciplines and is expanding to include new ones. This paper focuses on data mining techniques to retrieve valuable and applicable knowledge from broad datasets of under-studies utilizing LON-CAPA educational sources. The goal is to establish strategies that can provide knowledge which can easily be used by teachers to improve comprehension of undergraduates, To consider variations of school work problems, to coordinate instructional system designs all the more unmistakably, to anticipate agreements that under-study may pursue on various kinds of points, and to offer timely critique of under-study, etc. The specific field presented an overview of the LON-CAPA system, pointing out where we are going to use data mining techniques.

REFERENCES

- Minaei - Bidgoli, D. A. Kashy, G. Kortemeyer, and W. F. Punch (2003). "Predicting student performance: An application of data mining methods with an educational Web-based system", *33rd Annual Frontiers in Education (FIE 2003)*, Westminster.
- Al-Radaideh, Qasem A., Md Nasir Sulaiman, Mohd Hasan Selamat, and Hamidah Ibrahim (2005). "Feature selection by ordered rough set based feature weighting", *International Conference on Database and Expert Systems Applications*, pp. 105-112.
- Khan Z. N. (2005). "Scholastic achievement of higher secondary students in science stream", *Journal of Social Sciences*, Vol. 1, No. 2, pp. 84-87.
- Mazza, Riccardo and Christian Milani (2005). "Exploring usage analysis in learning systems: Gaining insights from visualizations", *Workshop on usage analysis in learning systems at 12th international conference on artificial intelligence in education*, pp. 65-72.
- Damez, Marc, Thanh Ha Dang, Christophe Marsala, and Bernadette Bouchon-Meunier (2005). "Fuzzy decision tree for user modeling from human-computer interactions", *Proceedings of the 5th International Conference on Human System Learning, ICHSL*, Vol. 5, pp. 287-302.
- Al-Radaideh .Q. A. E. W. Al-Shawakfa, and M. I. Al-Najjar (2006). "Mining student data using decision trees", *International Arab Conference on Information Technology (ACIT'2006)*, Yarmouk University, Jordan.
- Borstorff P. C., and S. K. Lowe (2007). "Student perceptions and opinions toward E-Learning in the college environment", *Academy of Educational Leadership Journal*, Vol. 11, No. 2, p. 13.
- Bray M. (2007). "The shadow education system: private tutoring and its implications for planners", (2nd ed.), UNESCO, PARIS, France.
- Merceron, Agathe, and Kalina Yacef (2005). "Tadaed for educational data mining", *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, Vol. 7, No. 1, pp. 267-287.
- Rossi P. (2009). "Learning environment with artificial intelligence elements", *Journal of E-Learning and Knowledge Society*, Vol. 5, No. 1, pp. 67-75.
- Alaa Mustafa (2009). "Mining Student Data to analyze Learning behavior", Research gate.
- Crist'obal Romero (2010). "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications And Reviews*, Vol. 40, No. 6.

Corresponding Author

Ritika*

Research Scholar, OPJS University, Churu, Rajasthan