

An approachable analytical study on Educational Data Mining

Karan Sukhija

Research Scholar,
Panjab University, Chandigarh, India.

Abstract - Data mining is the process of extracting evocative information and to develop significant association amid variables accumulated in data warehouse. It is a computational process of discovering patterns from huge databases and focuses on mining knowledge from the available data. Mining in educational environment is called educational data mining that is concerned with developing novel methods to discover knowledge from educational database. Educational data mining is an emerging research area that applies assorted tools and techniques of data mining to discover data in the field of education. It focuses on discover new patterns in captured data, developing novel algorithms or models for the escalation of education environment. This paper has fourfold objective. Firstly, it focuses on educational data mining as an emerging area with the combination of data mining and educational data. Secondly, it throws light on EDM process that applies data mining techniques to design educational systems. Thirdly, it considers all the EDM components to attain the educational objectives. Lastly, it concentrates on future research of EDM field viz. developing user-friendly mining tools, developing decision supports and recommendation engines and standardization of methods and data.

Key Words: (Educational Data, Data Mining (DM), Educational Data Mining (EDM), EDM Process, EDM Tools, Knowledge Discovery in Databases (KDD)).

1. INTRODUCTION

The encroachment in education domain leads to the development of gigantic information. With the increased volume of data, mining of knowledge from this huge data is the challenging task [1]. Data mining is a process applied to diversified fields to discover significant associations amid variables in huge data sets and is termed as Knowledge Discovery in Databases (KDD). Currently, data mining techniques are applied in the field of education to proficiently handle and extort undiscovered knowledge from the data [2]. Educational data mining is defined as follows: "Educational Data Mining is a promising discipline, concerned with developing methods for exploring the unique types of data gathered from educational settings, and applying those methods to better understand learner's behavior, and the environment which they learn in." [3]

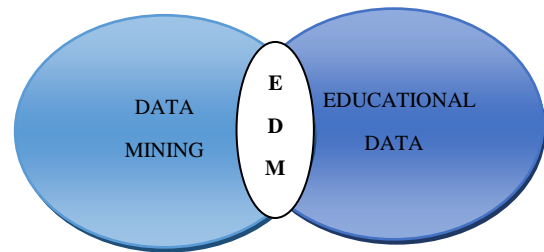


Fig.1. Educational Data Mining

EDM is an interdisciplinary field which inherits the properties from different domains such as Information retrieval, Psychometrics, artificial Intelligence, Domain driven, Machine learning, Learning analytics, Databases, Cognitive Psychology and so on [4]. The large repositories of data generated from different sources should be analysed to fulfil the goals in education. The foremost objective of EDM is discussed as below [5] [6]:

- **Learner Modeling** is concerned with the conception of learner models which includes learner behavior and their learning style, learner's performance and educational environment that helps in formation of their skills and solves their problems.
- **Domain Modeling** is concerned with the scheming of methods, tools and techniques applied for the improvement of particular branch or institution.
- **Learning System** focuses on the development of the system to learn and study the impact of educational support. E.g. Pedagogical support.
- **Computational Models** are devised that comprises of learner's modeling and domain modeling.
- **Impact of resources** is studied related to available resources viz. infrastructure, human resources, and Industry-academic relationship in the organization.

To congregate all the aforementioned objectives, a study of EDM is needed to deliver the quality education. The core objective of this manuscript is to offer brief knowledge about emerging research area i.e. educational data mining. The manuscript is organized as follows. In section 2 EDM Process is defined. Section 3 focuses on various modules of EDM. Section 4 highlights the future work and finally the paper is concluded in section 5.

II. EDM PROCESS

Educational data mining is an iterative and knowledge discovery process that has raw data (i.e. learner’s usage data, course information, academic data etc.) available from various educational environment viz. traditional classrooms, e-learning systems, adaptive and intelligent web-based educational systems as mentioned in figure 2 [7]. Further, this raw data is preprocessed and validated to discover associations among variables and data items. After data preprocessing a variety of data mining tools and techniques viz. prediction, clustering, classification, relationship mining, pattern matching etc. will be applied to get the results. The obtained result and interpretation will be offered to different users of education environment along with some suggested recommendations for problem-refinement [8].

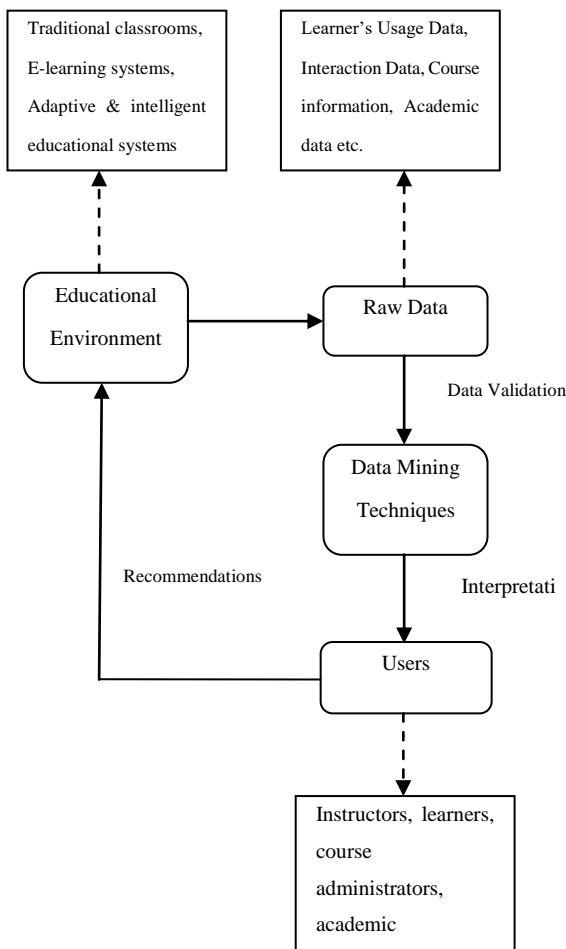


Fig.2. Applying data mining techniques to the design of educational systems

III. EDM MODULES

The major components of educational data mining are users and stake holders, EDM methods, applications of EDM, educational data, EDM tools and educational environment as discussed in figure 3 that will altogether applied to attain the objectives of EDM [9].

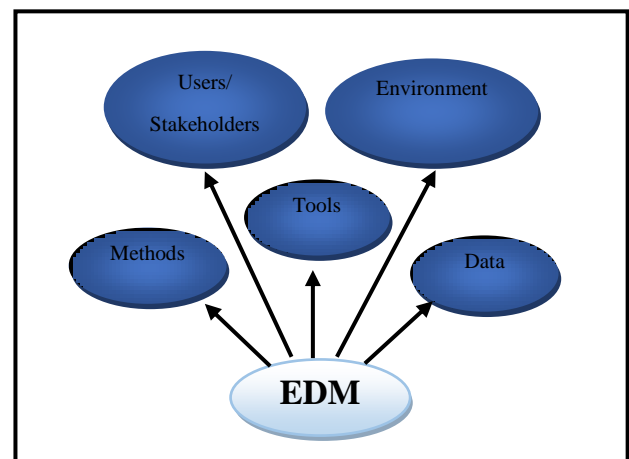


Fig.3 Modules of EDM

a. **Users and Stakeholders:** To attain the educational objectives, there are four classes of users and stakeholders as discussed below:

- **Learners:** The major task of learners is to provide feedback or recommendations related to educational environment that will helps in improvement of learning performance.
- **Educators:** Their purpose is to analyze learner’s behavior, learning environment, cognitive and social aspects that has impact on the teaching manner.
- **Educational Researchers:** Their task is to develop novel tools and techniques for the escalation of educational system.
- **Administrators:** Their responsibility is to utilize the resources available in the educational environment.

b. **EDM Methods:** It is one of the essential components of EDM that helps in relationship mining among variables, clustering of various subjects combination and prediction of learner’s behavior[7] [8]. These methods help in retrieval of web data from a variety of educational environments. Below mentioned are some of the data mining techniques used for educational data.

(i). **Prediction:** This technique is applied to analyze learner's performance, learner's behavior and their drop-out rate [10] [11]. It is used to derive predicted variable (single variable) from predictor variables (combination of variables). This technique is categorized as follows:

- *Classification* is used to predict class labels from (discrete or continue). Some popular classification methods include logistic regression, support vector machines and decision trees [12].
- *Regression* is used to predict from continuous variable. Some popular regression methods within educational data mining include linear regression, neural networks [13].

(ii). **Clustering:** It is an unsupervised technique applied for grouping objects into classes of similar objects [14]. In educational data mining, clustering technique is implemented to cluster learners according to their learning behavior [15].

(iii). **Relationship mining:** It is used to find out relationship among variables in a data set and form rules for specific purpose. Further, it is classified into four types:

- *Association rule mining* is used to identify association amid attributes in data set, mining interesting correlations, frequent patterns among data items for discovering learner's mistakes often stirring together while solving exercises [16] [17].
- *Correlation mining* is applied to discover linear correlations among variables i.e. positive or negative.
- *Sequential pattern mining* is used to find out sequential patterns i.e. the occurrence of one set of items is followed by another set of items [18]. It is based on temporal relationship between variables to predict which group a learner belongs to [19].
- *Casual data mining* is used to come across casual relationship among variables by analyzing the covariance of two events or by using information about how one of the events was trigger.

Some other EDM methods are distillation of data for human judgment used to present data in summarize and visualized way [20], discovery with models helps in relationship mining or prediction and knowledge tracing used to monitor student knowledge and skills over time [21].

c. **EDM Environment:** It is the learning environment offered by various educational institutions to the different users and stakeholders. EDM environment is be of following categories:

- *Traditional Class room Environment* is a formal environment where face-to-face communication takes place among users viz. schools and colleges where lectures are delivered by teacher to students in the classrooms [4].
- *Online/Web Based Environment* is an informal environment which works with the help of internet i.e. e-learning, Web Based performance prediction [12] and adaptive educational systems.
- *Computer Based Learning* is a hybrid environment of both formal and informal interaction. In this learning environment users can also work Offline viz. Intelligent Tutoring System (ITS) [22], learning Management system.

d. **EDM Data:** The huge amount of data collected from heterogeneous resources is used for decision making and learning process can be of any type as discussed below:

- *Private data* is concerned to data gathered from various academic institutions i.e. direct environment generates offline data.
- *Public data* is related to e-learning, web logs, e-mail, text data etc. i.e. indirect environment generates online data [22].

As discussed in the following figure 4, educational data viz. log files, learner's interaction data, and social network data types are estimated to grow in the future. This manuscript is leaning to the challenges involved with uncovering hidden patterns or extracting knowledge from large data sets by applying different data mining approaches and techniques as discussed further in table 1.

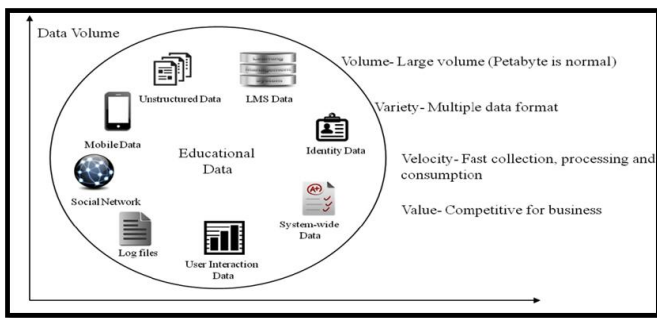


Fig. 4. Growth of different Educational Data

- a. **EDM Tools:** A variety of tools are discussed in table 1 for mining the repositories of data based on their features, techniques, and working environment [19]. Different tools will be used on the basis of the respective goal such as given in table 1.

Table 1. EDM Tools

Sr No.	Tools	Features	Technique
1	WEKA Tool	Provides machine learning algorithms for various data mining tasks and suitable for designing novel machine learning methods.	Data pre-processing, classification, Regression, clustering, association rules, and visualization.
2	Intelligent Miner	Provides firm integration with DB2 relational database system, and also offers scalability of Mining Algorithm	Association Mining, Classification, Regression, Predictive Modeling, Clustering, Sequential Pattern Analysis
3	MSSQL Server 2005	Provides data mining functions both in relational database system and Data Warehouse system environment.	Integrates the algorithms developed by third party vendors and application users.

4	Mine Set	Provides Robust Graphics tools viz. rule visualize, Tree visualizes, Map visualize and scatter visualizes	Association Mining, Classification, advanced statics and visualization tools
5	CART	Provides binary splitting and post pruning for Classification i.e. Decision Tree and for Prediction i.e. Regression Trees.	Classification, Regression Tree
6	ALPHA MINER	Provides the best cost and performance ratio for data mining applications.	Versatile data mining functions
7	Carrot	Provides ready-to-use components for fetching search results from various sources.	Clustering
8	Oracle Data Mining	Provides an embedded data warehouse infrastructure for multidimensional data analysis	Association Mining, Prediction, Classification, Regression, Clustering
9	SPSS Clementine	Provides an integrated data mining development environment for end users and developers	Association Mining, Clustering Classification, Prediction and Visualization tools.

Table 1 discusses various EDM tools used to mine educational data from heterogenous resources and further data mining techniques are applied to discover relationships among data items and pattern analysis of learner’s behaviour, clustering of learner’s records and classification of learning environments. The following

figure 5 depicts the analysis of data mining tools using over years from 1998 to 2012.

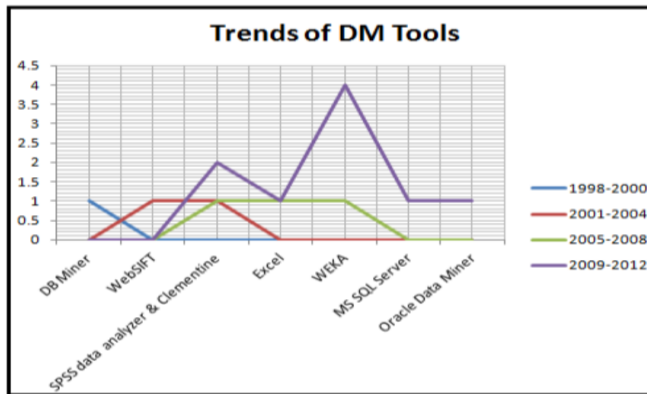


Fig.5. Trends of data mining tools [6]

The following figure 6 depicts the analysis of data mining techniques using over years from 2005 to 2012.

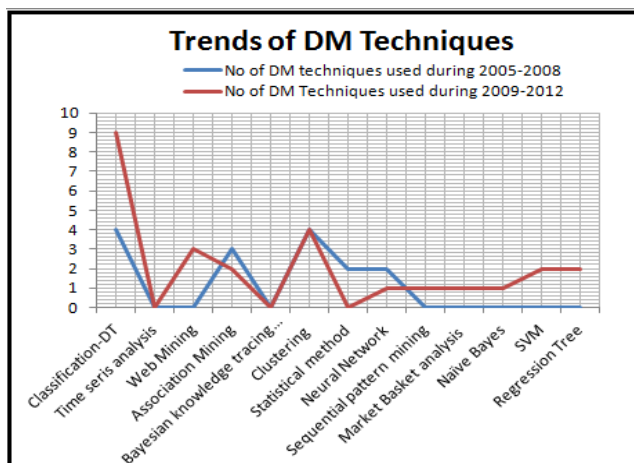


Fig.6. Trends of data mining techniques [6]

IV. FUTURE RESEARCH IN EDM

Educational data mining discover the hidden patterns from raw data to support learners, teachers, and institutions to understand and predict personal learning needs and performance to answer the educational queries.

- Development of user-friendly mining tools
- Developing decision supports and recommendation engines
- Standardization of methods and data
- Developing context-adapted models
- Integration with the e-learning system
- Specific data mining techniques

- Development of tools for protecting individual privacy.
- Developing more generalized tools that can be used by expert and non-expert user easily.

V. CONCLUSION

Educational Data Mining is relatively new and promising fields of research that aim to improve educational experiences by helping stakeholders (instructors, students, administrators and researchers) to make better decisions using data. It is a computational process of discovering patterns from heterogeneous databases and focuses on mining knowledge from the available data from various educational institutions. EDM growth has been boosted by increasing computer capacity to store and analyze huge amounts of data and the availability of statistical, machine-learning and data-mining methods and techniques. It has been evolved as multidisciplinary scientific learning area, rich in data, methods, tools and techniques used to provide better learning environment for educational users in educational context. Trends in data mining tools and techniques are discussed over the year from 1998 to 2012. This work focuses on research trends in offline, online and uncertain data, useful data sources, links etc in an educational context.

VI. REFERENCES

- [1] Xindong Wu, Xingquan Zhu, Gong-Qing Wu and Wei Ding, "Data Mining with Big Data". IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 1, 2014, 1041-4347.
- [2] Barahate Sachin R., Shelake Vijay M, "A Survey and Future Vision of Data mining in Educational Field", Second International Conference on Advanced Computing & Communication Technologies, 2012.
- [3] www.educationaldatamining.org.
- [4] Romero Cristobal & Ventura Sebastian, "Data mining in education", WIREs Data Mining Knowledge Discovery.
- [5] Baker and K. Yacef. "The State of Educational Data Mining in 2009: A Review and Future Visions." Journal of Educational Data Mining, 2009, 1 (1): 3-17.
- [6] C. Romero *, S. Ventura, "Educational data mining: A survey from 1995 to 2005", Expert Systems with Applications 33, 2007, 135-146.
- [7] Baker, R.S.J.D. in press. "Data Mining For Education", In International Encyclopedia of Education (3rd

- edition), B. MCGAW, PETERSON, P., BAKER Ed. Elsevier, Oxford, UK. 2009.
- [8] Ahmad A. kardan , Hamid Sadeghi, "A Decision Support System for Course Offering in Online Higher Education System" Int. Journal of Computational Intelligence Systems Vol. 6 No.5, Sep 2013, 928-942.
- [9] Jain, A. K., Murty, M. N., & Flynn, P. J., "Data clustering: A Review," ACM Computing Surveys, 31(3), 1999, (pp. 264– 323).
- [10] Amershi, S., Conati, C., "Automatic Recognition of Learner Groups in Exploratory Learning Environments," Proceedings of ITS 2006, 8th International Conference on Intelligent Tutoring Systems, 2006.
- [11] Dai Shangping , Zhang Ping." A Data Mining Algorithm In Distance Learning". IEEE, 2008.
- [12] Merceron, A., & Yacef, K., "Mining student data captured from a web-based tutoring tool: Initial exploration and results," Journal of Interactive Learning Research, 15(4), 2004, (pp. 319–346).
- [13] Agarwal, R., & Srikant, R., "Mining sequential patterns," In Proceedings of the eleventh international conference on data engineering, Taipei, Taiwan, 2005, (pp. 3–14).
- [14] Wang, W., Weng, J., Su, J., & Tseng, S., "Learning portfolio analysis and mining in SCORM compliant environment," In ASEE/ IEEE frontiers in education conference, 2004, (pp. 17–24).
- [15] Scheuer, O. & McLaren, B.M., "Educational Data Mining". In the Encyclopedia of the Sciences of Learning, 2011, Springer.
- [16] M. Bienkowski, M. Feng, and B. Means, "Enhancing teaching and learning through educational data mining and learning analytics: An issue brief," Washington, DC: SRI International, 2012.
- [17] R. Nisbet, J. Elder IV, and G. Miner, Handbook of statistical analysis and data mining applications: Access Online via Elsevier, 2009.
- [18] J.A. Lara, D. Lizcano, M.A. Martínez, J. Pazos, and T. Riera, "A System for Knowledge Discovery in E-Learning Environments within the European Higher Education Area- Application to student data from Open University of Madrid, UDIMA," Computers & Education, 2013.
- [19] D. Taniar, W. Rahayu, V.C.S. Lee, and O. Daly: "n Exception rules in association rule mining Applied Mathematics and Computation, 205(2): 735-750 (2008).
- [20] M.Z. Ashrafi, D. Taniar, and K.A. Smith: Redundant association rules reduction techniques, International Journal of Business Intelligence and Data Mining, 2(1): 29-63 (2007).
- [21] P. Williams, C. Soares, and J.E. Gilbert: A Clustering Rule Based Approach for Classification Problems. International Journal of Data Warehousing and Mining 8(1): 1-23 (2012).
- [22] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful knowledge from volumes of data," Communications of the ACM, vol. 39, pp. 27-34, 1996.