

A PROPOSED EDM FRAMEWORK FOR IMPROVING STUDENT PERFORMANCE

Mr. Bhavesh R. Patel

Abstract— Educational Data Mining (EDM) is an emergent discipline for exploring data, and a method to support learning and teaching processes. In this paper, we proposed Educational Intelligence (EI) Framework by combining BI technologies with various EDM algorithm techniques. In last decade, the higher education in India has grown manifold. Private participation in establishing new institutions, encouraged by the government, forced the higher education to revisit their scope and objectives in the long run to sustain. This paper aim is to improve the efficiency of higher educational institutions by applying data mining techniques such as clustering, decision tree, association etc. This paper also describe that how data mining algorithms can be applied to higher education processes for enhancing student's performance. From the analysis of this framework, the groups of students who have excellent skills or vice versa can be identified. It also optimizes the time to perform current and historical data analysis. The weaknesses and strengths of the student can also be obtained. Finally, students' future potential areas of studies can be predicted using the framework.

Keywords— Business Intelligent, Educational Intelligence, Educational Data Mining, Educational Data Warehouse Introduction, Data Mining Algorithms, Business Intelligent.

I. INTRODUCTION

Nowadays, higher educational organizations are placing in a competitive environment and are aiming to get more competitive advantages over the other competitors. To remain competitive in educational domain these organizations need deep and enough knowledge for a better assessment, evaluation, planning, and decision-making. Most of the required knowledge can be extracted from the historical and operational data that reside in the educational organization's database. Data mining techniques are analysis tool that can be used to extract meaningful knowledge from large data sets. [1]

Currently the concepts and techniques of data mining is a promising frontier in database systems and new database system applications. Data mining is considered as a highly desirable tool being applied to many application areas such as finance, banking, telecommunication, medicine, retail industry and education [2, 3, 4, 5]. One of these application domains is higher educational system. Data mining extracts previously unknown, valid, novel potentially useful and understandable patterns in educational data of large databases[6]. The

discovered hidden patterns enable the higher educational system in making better decisions and having more advanced plan in directing students. As a result the higher educational systems, which are delivered through universities, post-academic study schools and university collages can be facilitated in having more accurate, efficient and effective educational processes. Our main contribution in this paper is addressing the capabilities and strengths of data mining technology in the context of higher educational system. Higher educational system can enhance their educational processes according to our proposed analysis model.

II. CONCEPT OF DATA MINING

Data mining is the process of extracting useful knowledge and information including, patterns, associations, changes, anomalies and significant structures from a great deal of data stored in databases, data warehouses, or other information repositories . It is considered as one of the knowledge discovery step, which interacts with user or knowledge base. The interesting patterns are presented to users, and maybe stored as new knowledge in the knowledge base. Even though it is only one step in the entire knowledge discovery processes but an essential one since it uncovers hidden patterns for evaluation [6]. The data mining delivered a series of functionalities such as outlier analysis, evolution analysis, association analysis, classification, clustering and prediction [1, 6].

Data mining has been applied into many application domains such as biomedical and DNA analysis , retail industry and marketing, telecommunications, web Mining, computer auditing ,banking, fraud detection, financial industry, medicine, and education. Our survey on the current works in data mining field shows that one of the application domains that can take advantage of data mining benefits is education and this field needs much more attention[6,7].

Process of Data Mining [6]:

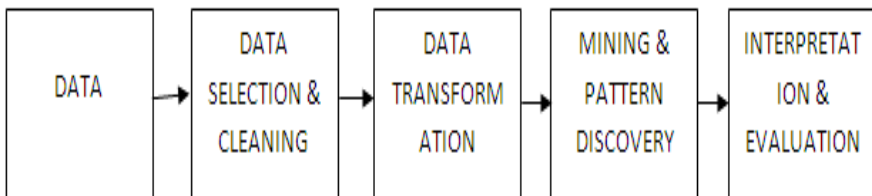


Fig. 1 Process of Data Mining

On what type of Data can we applied mining techniques?

- Different type of data can be used as the data source of data mining, such as:
- Relational
- Data warehouses
- Transactional databases
- Advanced DB and information repositories
- Object-oriented and object-relational
- Spatial databases
- Time-series data and temporal data [11]

III. DATA MINING IN HIGHER EDUCATIONAL SYSTEM

Data mining is considered as the most suited technology appropriate in giving additional insight into the lecturer, student, alumni, and other educational staff behavior and acting as an active automated assistant in helping them for making better decisions on their educational activities. As discussed before, lack of deep and enough knowledge in higher educational system may prevent system management to achieve their quality objectives.

Data mining technology can help bridging this knowledge gaps in higher educational system. Therefore the hidden patterns, association and anomalies, which are discovered by some data mining techniques, can be used to improve the effectiveness, efficiency and the speed of the processes. As a result, this improvement may bring a lot of advantages to the higher educational system such as maximizing educational system efficiency, decreasing student's drop-out rate, increasing student's promotion rate, increasing student's retention rate, increasing student's transition rate, increasing educational improvement ratio, increasing student's success, increasing student's learning outcome, and reducing the cost of system processes.[8,9,10]

In order to achieve the above quality improvement, we need a data mining system that can provide the needed knowledge and insights for the decision makers in the higher educational system. We have analyzed such a system and in the next section we will present and describe an analysis model for this system.

IV. BUSINESS INTELLIGENCE VS. EDUCATIONAL INTELLIGENCE

BI refers to skills, processes, technologies, applications and practices used to support decision making. BI software leverages on investment in data and systems to provide easy-to-use information that improves decision-making in an organization. There are many BI tools that provide several applications to suit certain purposes. Some applications produce an Executive Reporting System (ERS) while others require dashboard, graphical reports and even more powerful analytic programs that allow user to have more manipulations and flexibility over a set of data from DW.

Common functions of BI technologies are reporting, dashboard, OLAP, analytic, data mining, business performance management, benchmarking, text mining, and predictive analytics. BI often aims to support better business decision-making. Dashboard is multilayered performance management systems, built on BI and data integration infrastructures, enable organizations to measure, monitor and manage business activity using both financial and non-financial measures. Whereas scoreboards tend to chart the progress of tactical and strategic goals. Dashboards are often used to display and monitor Key Performances Indicator (KPI). Another main feature of BI technologies is the reporting tool. The reporting provided by BI tools is usually able to customize to the customers' needs [13].

In this paper, we are trying to apply the same concept for educational data. Several features that EI needs to have are[13,14]:

- A.** The ability of various data from different sources to be integrated in one large pool of database/Educational Data Warehouse (EDW).

The implementation of EDW will improve the process of handling large volume of data and rapid changes in educational data. Designing a physical design of database in multidimensional format is improving the query taken for producing reports. Furthermore, applying cube for OLAP engine provides efficiency and flexibility in educational data reports.

- B.** The ability to perform multiple analytic processes whether descriptive analysis or predictive analysis

The main importance of EDM is the ability to provide comprehensive analysis on students' data. Lecturers or teachers may understand students' individual behavior and could devise strategies to improve their achievements in learning process. The analysis provided by EI whether descriptive or predictive will help improving learning and teaching process. In traditional ways, it is difficult for educators to know each of their students. It is because some of them are having more than 100

students at certain times. Therefore, the ability to produce meaningful reports is important in learning process.

C. A user-friendly Graphical User Interface (GUI) for EI reports

Not all teachers are a computer literate but usually most of them have basic knowledge on computers. In order to ensure EI technologies will be supported by teachers, a simple and efficient GUI needs to be provided by EI reports. Teachers can customize and use the reports with minimal technical assistance.

V. RELATED STUDIES IN EDUCATIONAL INTELLIGENCE:

Educational Intelligent using data mining has emerged as an independent research area in recent years, culminating in 2008 with the establishment of the annual International Conference on Educational Data Mining, and the Journal of Educational Data Mining. Romero and Ventura [18] provides a comprehensive study of EDM from 1995 to 2005. It describes the need for analyzing the student data which can be used by students, educators and administrators. Galit [15] developed a system to warn weak students. Han and Kamber [06] discovered relationship among data. Walters and Soyibo [16] discovered relationship between academic performance and nature of their schools. Z.N. Khan [17] found Girls with high socio-economic status were relatively higher achievers in science stream and boys with low socio-economic status were relatively higher achievers in general. Hijazi and Naqvi, [17] using regression found factors like mother's education and student's family income were highly correlated with the student academic performance. Moriana et al. [18] used Analysis of variance (ANOVA) and it was observed that group involved in activities outside the school yielded better academic performance. Cortez and Silva [19] found Decision Tree and Neural Networks in some areas give same accuracy.

VI. DIFFERENT FRAMEWORK FOR EDUCATIONAL INTELLIGENT

MUSTAS FRAMEWORK:

The Multidimensional Students Assessment (MUSTAS) framework is a novel model, which consist of demographic factors, academic performance of the student and dimensional factors. The dimensional factors has further sub divided into three dimensions respectively self-assessment, institutional assessment and external assessment. The main objective of this framework is to identify the contribution of selected dimensions over academic performance of the student, which helps to teachers, parents and management about the student's pattern.

Understanding of the pattern may facilitate to redefine the education method, additional care on weakness, and promoting their abilities. The academic performance shows the present ability of the student and the demographic factors shows his personal lifestyle. Construction of this framework strongly believes each aspect considered for this framework is closely related to one another.[20]

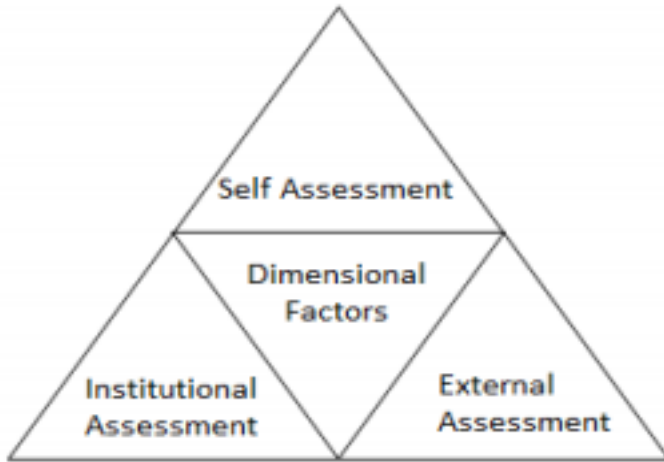


Fig. 2 Assessment Criteria

[source: IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011
ISSN (Online): 1694-0814]

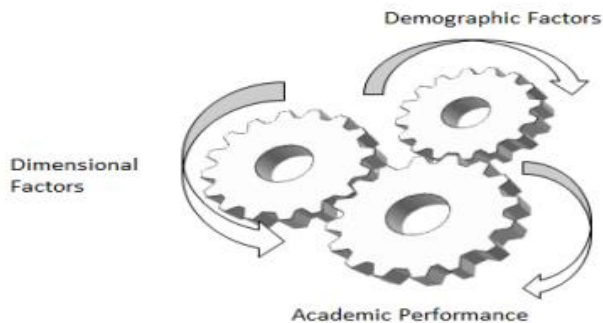


Fig. 3 Dimensional Factors

[Source: IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011
ISSN (Online): 1694-0814]

The dimensional factor helps to measure the student’s attitude. Self-assessment is measured through five questions, which express their personal interest towards studies. Institutional assessment is specially designed for lecturers/faculty and institution’s support towards studies. The third dimension is external assessment, which is designed to measure an external attribute contribution towards their studies.[20]

CHAID (CHI-SQUARED AUTOMATIC INTERACTION DETECTION ANALYSIS) BASED PREDICTION FRAMEWORK:

CHAID based Performance Prediction framework in Educational Intelligence was analysed by Ramaswamy, [21] and the results have proved to be accurate when compared to some other models in terms of accuracy in prediction

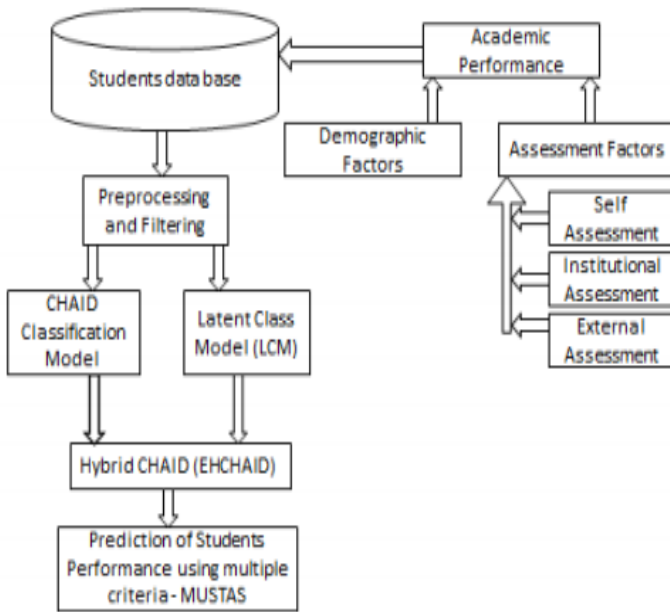


Fig. 4 Proposed Framework

. One limitation of CHAID is that segments are defined based on a single criterion variable. Given situations where multiple criteria exist, it is not clear how one should go about obtaining a single common segmentation. Using one dependent variable as the criterion may result in one set of segments, while use of an alternative dependent variable will likely yield a different set of segments.

Moreover, the categories of a predictor may merge in different ways depending upon which dependent variable is used, again leading to different segments.

[CHAID (CHI-SQUARED AUTOMATIC INTERACTION DETECTION ANALYSIS) BASED PREDICTION FRAMEWORK:] PROPOSED FRAMEWORK:

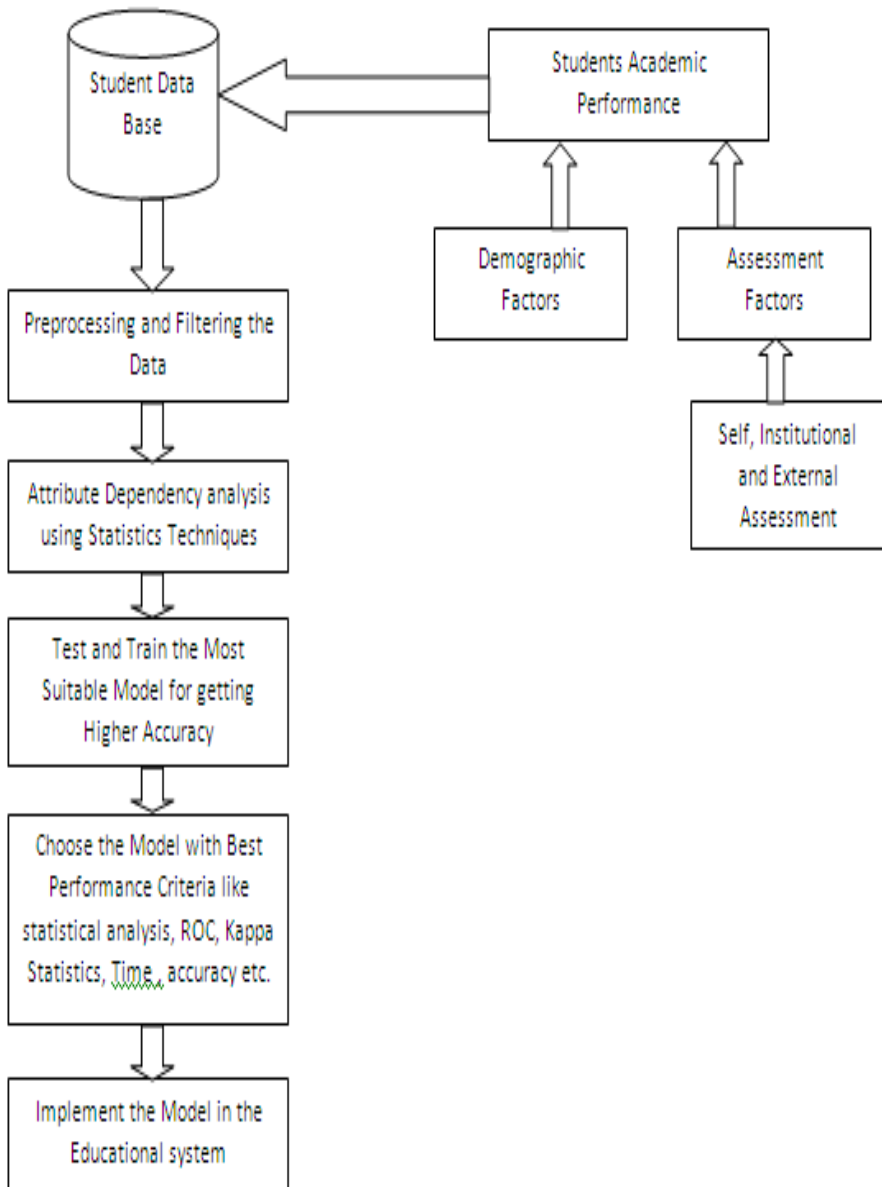


Fig. 5[Proposed Framework for Accuracy based Educational Intelligent]

As shown in figure we propose a generalized design framework for Educational Intelligence. Here history data is collected and data cleaning is done. After that attribute dependency analysis using chi square statistical analysis is done to clearly understand which all attributes have importance in decision making. Once this is properly understood attribute reduction and porting to model dependent data formats takes place.

For Weka based data mining projects we have to port it to arff based formats, for MATLAB based projects we have to port it to CSV/XLS formats and so on. Then the data mining models are built using training data and are tested using test data to decide which model performs in the best way for this domain. For those comparisons various statistical measures like ROC area, Kappa statistic, True positive, accuracy etc may be used. Once the best model is chosen that model can be implemented for actual predictions for future data.

VII. CONCLUSION

This paper describe that how can we get the educational intelligence by applying data mining techniques and Business Intelligence in to the higher education processes for enhancing student's performance. For that in this paper described a new proposed framework, by analyzing this framework, the groups of students who have excellent skills or vice versa can be identified. It also optimizes the time to perform current and historical data analysis. The weaknesses and strengths of the student can also be obtained. As a output of this framework we can also find out the best model for the accurate predicted result. We can also find out the attribute dependency by applying this proposed framework. To develop this proposed framework as a reference we have referred frameworks like **MUSTAS framework**, CHAID based framework and based on these we have proposed a new framework. As a feature work we can implement this framework by applying any data mining or Business Intelligent tools like Weka, Clamentine, Cognos, SAS, Business Object, Hyperion and Open Source Data Miner or BI/ETL Tools.

REFERENCES

- [1] N. Delavari, M. R. A. Shirazi, M. R. Beikzadeh. A New Model for Using Data Mining in Higher Educational System, 5th International Conference on Information Technology based Higher Education and Training: ITEHT '04, Istanbul, Turkey, 31st May-2nd Jun 2004.
- [2] J. Han, How can Data Mining Help Bio-Data Analysis. BLOKDD02: Workshop on data mining in Bioinformatics, 2002.

- [3] R. Feldman, Mining the Biomedical Literature using Semantic Analysis and Neural Language Processing Techniques, a link analysis approaches. Clear Forest Corporation, New York, 2003.
- [4] H. Edelstein, Building Profitable Customer Relationships with Data Mining", Two Crows Corporation, SPSS white paper-executive briefing, 2000.
- [5] W. H. T. Chang and Y. H. Lee, Telecommunications Data Mining for Target Marketing, Journal of Computers, Vol. 12, No. 4, December 2000, pp.60-74.
- [6] J. Hans and M. Kamber. Data Mining: Concepts and Techniques. Simon Fraser University, Morgan Kaufmann publishers, ISBN 1-55860-489-8. 2001.
- [7] M. Reza and N. Delavari, Processes in Higher Education System, TS3B-2 M2 USIC 2004,
- [8] Information Technology journal 5(3): ISSN 1812-5638, Data mining for better higher education system.
- [9] Journal of Data Science 8(2010) Data mining approach for Identifying Predictors of Student Retention from sophomore to Junior Year.
- [10] Ryan S.J.D. Baker, The State of Educational Data Mining in 2009: A Review and Future Visions [pp]
- [11] Academic Analytics and *Data Mining* in Higher Education www.academics.georgiasouthern.edu/ijstot/v4n2/.../PDFs/_BaeplerMurdoch.pdf.
- [12] Aziz, W. Rizhan and H. Hassan "Intelligent System for Personalizing Students' Academic Behaviors" International Journal on New Computer Architectures and Their Applications (IJNCAA) 2(1): 138-153 The Society of Digital Information and Wireless Communications, 2012 (ISSN: 2220-9085)
- [13] Cognos, The right architecture for business intelligence, CognosInc: Burlington, USA, 2008.
- [14] Jason, O., W., K., Kannan, S., and Asirvadam, V., S.: Data Mining and Warehousing Approaches on School Smart System: A Conceptual Framework,

In Proceedings of Knowledge Management International Conference, pp. 20-24 (2008).

- [15] Galit.et.al. 2007."Examining online learning processes based on log files analysis" : a case study. Research, Reflection and Innovations in Integrating ICT in Education.
- [16] Madhyastha.T.and Tanimoto, S., 2009." Student Consistency and Implications for Feedback in Online Assessment Systems. "In Proceedings of the 2nd International Conference on Educational Data Mining, pp81-90.
- [17] Z. N. Khan, Scholastic Achievement of Higher Secondary Students in Science Stream, Journal of Social Sciences, Vol. 1, No. 2, 2005, pp84-87.
- [18] J. A. Moriana, F. Alos, R. Alcalá, M. J. Pino, J. Herruzo, and R. Ruiz, Extra-Curricular Activities and Academic Performance in Secondary Students, Electronic Journal of Research in Educational Psychology, Vol. 4, No. 1, 2006, pp35-46.
- [19] P. Cortez, and A. Silva, Using Data Mining To Predict Secondary School Student Performance, In EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp5-12.
- [20] G.Paul Suthan and Dr. Santhosh Baboo Hybrid CHAID a key for MUSTAS Framework in Educational Data Mining, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814]
- [21] M. Ramaswami and R. Bhaskaran. A CHAID Based Performance Prediction Model in Educational Data Mining, Madurai Kamaraj University, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1, January 2010

AUTHORS' PROFILE



Prof. Bhavesh Patel,

Ph.d (Pursuing), M.Phil (CS), M.C.A.

He is working as assistant professor in Ganpat University, M.C.A Department. Area of Interest is Data Mining, Software Engineering. Presented four National levels Paper. He has ten year teaching experience in the field of academic.