International Journal of Computing Science and Information Technology, 2018
ISSN: 2278-9669, Aprily 2018 (http://ijcsit.org)

Review of EDM for Higher Education Sustainability

N Kartik[1], Dr K A Venkatesh[2]
[1]Research Scholar
Presidency University
Bangalore
[2]Professor
Presidency University
Bangalore
[1]*nkartik.mca@gmail.com*, [2]*ka.venkatesh@presidencyuniversity.in*

Abstract

Educational Data Mining (EDM) is a rising field investigating data in the educational setting by applying diverse Data Mining (DM) techniques and tools. It provides vital information for various stakeholders involved in Educational Data Mining. Lot of work has been done in forecasting weather and predicting the share value but a very little work has been done in the field of education which is one of the important areas for the development of the nation. This paper is a systematic review of EDM for higher education sustainability.

Keywords: Learning Management System (LMS), AICTE

1. Introduction

In last 20 years number of university and colleges have grown rapidly both in private sectors and public sectors, so they have a tough competition to survive in this education sector. Recently AICTE has asked to close down about 800 colleges across India as there are no takers for their seats. This becomes the alarming thing to other colleges. Quality education is one of the major factors to survive. Most of the information is digitally available now the biggest challenge how we use these data efficiently? [1] Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. While the analysis of educational data is not itself a new practice, recent advances in educational technology, including the increase in computing power and the ability to log fine-grained data about students' use of a computer-based learning environment, have led to an increased interest in developing techniques for analyzing the large amounts of data generated in educational settings. Lot of work has been done in the field of weather forecasting and prediction of shares price and we are able to reach an accurate result. Similarly, we can apply in the field of education as Education plays a vital role in the development of the nation. With Educational Data Mining we can find the hidden facts which are available. There are various stakeholders of Educational Data Mining like Students, Teachers, College, Parents and Non-Teaching staff.

2. EDM Applications

Over the years EDM is being applied in various goals. According to [9] its

a) Communicating to stakeholders: with an aim to communicate to the different stakeholders at right time to take necessary action accordingly.
b) Maintaining and improving courses: to help administrators and educators in determining how to improve the course with respect to its materials.
c) Generating recommendation: based on the activity in learning management system (LMS) to give appropriate suggestions. The most frequently used techniques for this type of goal are association, sequencing, classification, and clustering.
d) Predicting student grades and learning outcomes: to predict a student's final grades and outcome based on data from course activities like test, assignments, attendance etc.
e) Student Modeling: All students are not the same students can be modeled based on the characteristics such as satisfaction, motivation, learning progress or certain types of problems that negatively impact their learning outcomes.

3. Literature Survey

[2] One of the new student selection process usually used by the university, based on student grade from high school or vocational high school. In this selection process, processing involves large amounts of data. With the application of clustering algorithm, we can get an insight into the data and a huge amount of time can be saved. A similar technique can be applied in the selection of students for college who are not only good in studies but also in other fields like sports, cultural, extra co-curricular activities so that college will not only be producing good graduates but also be on top in other fields. Most of the time students spend on social networking site like Facebook, Twitter. [3] Student's learning experiences are taken from Twitter i.e., the emotions of the students and then analyze them to make decisions about the problems faced by students. A lot of unstructured data is present on Twitter during pre-exams we can capture the twits and capture in insight from the data. Also in LMS, we can capture the students chat, questions they are asking in the forum etc. Classification is a data mining technique that assigns categories to a collection of data in order to aid in more accurate predictions and analysis. An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent. [4]Author used classification technique to predict the student result (pass/fail) and association for finding relations in the subject marks. At a very early stage considering the association rules students can be asked to concentrate on the weak subjects which may affect them in the future semester for example, C programming in a lower semester may have an impact on learning and passing C++. A decision tree is a representation made out of nodes and arcs where an internal node presents a decision based on attribute values, and the arcs represent the choice made in the node. It ends on a leaf node, which represents the label or the class to be assigned. To classify a record with a decision tree, it starts by the root node and goes down one level at a time depending on the results of the conditions tested on every node; when it ends on a leaf node, the record is classified according to the label of that leaf node. [5]Using Initial Academic Information, Demographic and Socio-Economic and Academic Potential a model was developed which was detecting the students at risk. [6] a prediction model is proposed which helps the students to choose a course based on the type of data or information that he/she furnishes. The system which predicts the best course suitable for a student so that he will outperform in the selected course. In this dataset like Gender, Category, Sector, Branch is taken into consideration. [7]Canberra Distance as a similarity measure and cluster the students based upon their marks obtained in CAT I, CAT II and Quiz. Students are categorized into different clusters also in this way we can categorize unusual behavior or failure. With this higher category student can be given advanced training, failure student can be trained to pass and give the confidence to face examination and unusual students can be handled in a proper way and required action can be taken as early as possible. [8] The fluctuations in internal assessment marks decreases with each passing semester. Also attendance and student aggregate performance and that the performance decreases with decreasing attendance. We could also do clustering based on marks scored in each subject to get a greater insight of data.

Author name	Problem	Technique	Algorithms	Dataset
[2]I. M. Suartana et al	Analysis of New Student Selection using Clustering Algorithms	Clustering	K Mean	high school final exam grade, high school main subject grade, high school report grade, academic excellence
M. [3] B. Talawar et al	Analyzing Social Media Data in Educational Sectors Using Data Mining Techniques	Classification	Naïve Bayes	Unstructured text data
S. A. [4] Kumar	Edifice an Educational Framework using Educational Data Mining and Visual Analytics	Classification and Association rule	C4.5 decision tree algorithm	Attendance, Marks of each subject, gender
C. E. [5] Lopez Guarin et.al	A Model to Predict Low Academic Performance at a Specific Enrollment	Classification	Naïve Bayes decision tree	Initial Academic Information Demographic and Socio-Economic

	Using Data Mining			Academic Potential
[6] Suganthi, G	Predicting Employability of Students Using Data Mining Approach	Clustering	X Mean Algorithm, Support Vector Clustering	Gender, Category, Sector, Branch
[7] K. N. Shah et.al	Clustering students' based on previous academic performance	Clustering	Euclidean distance, Manhatta distance, Squared Euclidean distance, Canberra distance	Previous academic marks, quiz score
[8]Bindiya M Varghese	Clustering Student Data to Characterize Performance Patterns	Clustering	K-Means Clustering Fuzzy C-Means	Attendance Internal mark assessment Seminar assessment Class assignment assessment University marks scored

4. Discussion

A lot of scopes are there for:

- Blended EDM tool which monitors traditional as well as online learning
- EDM to identify student skills, train them on that and make them employable
- Detecting Malpractice in online examination

As discussed above authors have discussed various applications of data mining in the field of education, we can apply this in the educational system to improve the quality of education.

Still, a lot of work is being carried out to improvise the existing EDM algorithms

5. Conclusion

There is very little research conducted in India. Institutions give major focus on infrastructure, qualified faculty, marketing of institutions, Value-added programs etc. EDM techniques should also be implemented for better decision making by management and also by doing so, we can understand student's trend in the better way so that it can be applied to upcoming batches.

6. Acknowledgement

I would like to thank Dr K A Venkatesh, for providing me inputs and feedback that greatly improved the manuscript.

REFERENCES

[1] R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009 : A Review and Future Visions," J. Educ. Data Min., vol. 1, no. 1, pp. 3–16, 2009.

[2] I. M. Suartana and A. I. N. Hidayat, "Analysis of New Student Selection using Clustering Algorithms," IOP Conf. Ser. Mater. Sci. Eng., vol. 288, p. 12079, 2018.

[3] M. B. Talawar and N. C. Kundur, "Analyzing Social Media Data in Educational Sectors Using Data Mining Techniques," vol. 7, no. 4, pp. 2125–2129, 2016.

[4] S. A. Kumar, "Edifice an Educational Framework using Educational Data Mining and Visual Analytics," Int. J. Educ. Manag. Eng., vol. 6, no. 2, pp. 24–30, 2016.

[5] C. E. Lopez Guarin, E. L. Guzman, and F. A. Gonzalez, "A Model to Predict Low Academic Performance at a Specific Enrollment Using Data Mining," Rev. Iberoam. Tecnol. del Aprendiz., vol. 10, no. 3, pp. 119–125, 2015.

[6] Suganthi, G., Predicting Employability of Students Using Data Mining Approach, International Journal of Information Research and Review(2017)

[7] K. N. Shah, S. Kothuru, and S. Vairamuthu, "Clustering students ' based on previous academic performance," vol. 3, no. 3, pp. 935–939, 2013.

[8] J. Varghese, Bindiya M; Tomy, Jose; Unnikrishnan, A; Poulose, Clustering Student Data to Characterize Performance Patterns. International Journal of Advanced Computer Science and Applications, 138–140 (2010).

[9] C. Romero, S. Ventura, M. Pechenizkiy, R. Baker, Handbook of Educational Data Mining Taylor & Francis 2010.