ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Vol 9, Issue 7, July 2022

Improving Data Quality for Educational Data Mining (EDM) for Indian Ed-Tech Start-Ups

[1] Tamal Mukherjee

[1] PhD Scholar, Symbiosis Center for Research & Innovation, Pune, India.
Corresponding Author Email: [1] tkm.scit@gmail.com

Abstract*— Educational Data Mining (EDM) is one of the most widely used data mining application areas. Improving the process of EDM is a challenge for Indian Ed Tech Start-ups. There exist a lot of efficient statistical algorithms for EDM. But for successful application of these high-tech statistical methods, the back-end data fetching process should be properly in place. That is only possible if the back-end data quality of all these EDM data are maintained and managed efficiently and proficiently. The technical implementation of this remains a huge challenge. In this particular paper, we aim at taking care of this issue by improving the back-end data management system quality which acts as the backbone of any EDM procedure.

Index Terms*— Educational Data Mining (EDM), Back-end requirements, MIS in EDM, Root Cause Analysis, Preventive Model, Ed-Tech, Start-up

I. INTRODUCTION

There are various methods using which EDM can be done. Though the methodology followed by each of these methods differ from one technique to another, the underlying purpose converges to the same concept – understanding the student community better than ever through the proper usage of the existing data store. Though the scope of EDM is not limited to any specific level of educational institute but presently it is in practice mostly at the university level. Keeping that in mind we are also, throughout this particular research paper, attempting to touch upon the problem areas of EDM for Indian Ed-Tech start-ups.

II. OBJECTIVES

i. Figuring out Baker's models of Educational Data Mining (EDM).
ii. Understanding the basic Front-End and Back-End requirements of EDM for Indian Ed-Tech start-ups.
iii. Carrying out a field survey to find out the biggest challenge(s) of the above back end data store maintenance for Indian Ed-Tech start-ups.
iv. Doing the Root Cause Analysis (RCA) of all the biggest challenge(s).
v. Coming up with a Model to tackle the root cause(s) obtained after RCA.

III. LITERATURE REVIEW

Data mining is one of the fastest emerging trends in the modern technical world. This is used by different sectors to make good use and sense out of the hugely available repository of data existing with them. Educational sector too is no exception. This sector is using data mining in the name of Educational Data Mining (EDM). The EDM community website [1] defines educational data mining as: "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."

Baker's five methods are commonly used in extensive educational data mining. His models of EDM follow different statistical methods for visualization – Prediction, Clustering, Relationship Mining, and Distillation of data for human judgment and Discovery with Models. In these models, estimates are valued upon the observed data following some underlying statistical functions. Though his work at this front was mostly a factual extension of Moore's work on Statistical Data Mining, Baker's models led to some revolutionary changes in the EDM domain of research and development.

To implement any of the above methods for the successful usage of EDM, there are two basic requirements one system needs to comply with - The front end requirements and the back end requirements of the system. To maintain the data where the volume is big, the best possible solution is Management Information System (MIS). But maintaining data through an MIS suitably for the purpose of EDM is a huge challenge. In the rest of the paper, we will try to find out a solution to this problem so that the back-end data can be maintained and managed to the fullest.

Different specialists defined MIS [2] differently. Let us have a look at some of those. According to Davis and Olson, "MIS is an integrated user-machine system for providing information to support operations, management and decision-making functions in an organization. The system utilizes computer hardware and software manual procedures/models for analysis, planning, control and decision-making and a database." Kelly defined MIS as "a combination of human and computer based resources which result in collection, storage, retrieval, communication and use of data for the purpose of efficient management of operations and for Business Planning." So, by combining these

ISSN (Online) 2456 -1304
International Journal of Science, Engineering and Management (IJSEM)
Vol 9, Issue 7, July 2022

definitions we can say that MIS is basically an integrated system which transforms the data (inputs) into reports (outputs) for facilitating decision-making through processing using various components like Hardware, Software, Database, Procedures and Personnel. MIS can be further classified into four different models as under:-

a) Decision Support System (DSS)
b) Executive Support System (ESS)
c) Expert System (ES)
d) Artificial Intelligence (AI)

IV. RESEARCH METHODOLOGY

Objectives of the research

- To find out the various challenges involved in all the above four practices of MIS (i.e. Decision Support System, Executive Support System, Expert System and Artificial Intelligence) through field survey.
- To comment on the biggest challenge (s) corresponding to each of those four practices of MIS (i.e. Decision Support System, Executive Support System, Expert System and Artificial Intelligence) through that very field survey so that it can be removed to increase the efficiency of EDM.

Sampling

The basic information related to Decision Support System, Executive Support System, Expert System and Artificial Intelligence was gathered via literature review. Then, we aimed at a group of people who either worked in the domain of MIS or are working in the domain of MIS as a part of EDM.

Data Collection

It was aimed to reach out to more than 120 professionals but was able to collect the data only from 70 respondents. This survey was conducted during the months of May, 2022 to June, 2022 via the use of electronic mails and telephonic interviews.

Finding the biggest challenge(s)

For Decision Support System, 'History Maintenance' of data is found to be the biggest challenge.

For Executive Support System, 'Drilling Down' of data is found to be the biggest challenge.

For Expert System, 'Archive Maintenance' of data is found to be the biggest challenge.

For 'Artificial Intelligence based System', 'Data Retrieval' is found to be the biggest challenge.

We simulated the concept of focus group over telephone to carry out a 'Why-Why Analysis'.

As a conclusion we found out that the root causes of all the respective biggest challenges of the four practices of MIS, converge to a single cause; which is – 'Data Loss at the time of Data Migration from one version to the next of the concerned system'. So, if this root cause can be eliminated MIS can resolve all the back-end issues related to EDM.

V. RECOMMENDATION

In order to make better use of all the changed forms of MIS in EDM, we must try and eradicate the root cause mentioned above. In order to do so, a data auditing model is suggested capturing the table level and element level changes during any data migration activity.

VI. CONCLUSION

So, for better and efficient Educational Data Mining, using all the variations of the MIS, according to this research the Indian Ed-Tech start-ups must start using the data-auditing model.

REFERENCES

[1] Moore, Andrew. [Online]. Available: http://www.educational datamining.org
[2] Oke, Jayant. 2000. 'Management Information System', pp. 1.4-1.21.