APPLICATION OF EDUCATIONAL DATA MINING IN NIGERIA: OBJECTIVES, METHODS, APPLICATIONS AND STATISTICAL SOFTWARE TOOLS

Author's name: AfeezTundeJinadu

Ph.D., Centre for Educational Research and Management Ibadan, Nigeria
E-mail: afeezjinadu95@gmail.com

Abstract	*Educational Data Mining (EDM) became known field to explore the data from educational point of views. This makes knowledge about transmitting the education, easier to enhance the quality of teaching and learning through careful planning and organisation of data involved. Careful planning and organisation can provide the desired education. Therefore, this paper presents the objectives, methods, applications and some of the statistical software tools for educational data mining.*
Keywords	*Educational Data Mining, Methods, Applications and Tools*

INTRODUCTION

The idea behind data mining on its own is to unravel the existing trends in data and create new and effective insights. Data mining is otherwise known as Knowledge Discovery in Databases (KDD) or sometimes Big Data. More specifically and applied to education, Education Data Mining-EDM involve systematic processing of large amount of educational data by figuring out salient latent facts and orders. Data come from several sources, such as data from parenteducational background, students background, teacher background, government policies, classroom environments, educational software, online courseware, tests scores among others. These sources increasingly provide large amounts of data, which can be analyses to easily address issues that may not previously feasible.

Education data mining focuses on the development of exploration methods of specific types of data coming from the educational context which are used to gain insight into manner in which students learn and identify those aspects that can improve learning and other educational aspects [1]. It then follows that information derived from learning situations are collated, arranged, sorted and important latent information are selected for gaining insights into later occurrence of such information for decision making purposes in Nigeria as it has been useful in other countries of the world.

Educational data mining helps with future pacing and provides a vision. If we look at the recent past; the education sector has seen an extreme vertical as well as horizontal growth. This kind of enormous growth brings along an elevation in educational data that are more of facts and figures[2]. However, helping educational institutions to improve the educational settings goes beyond generating facts and figures such as total number of students and teachers based on gender, location, schools and so on without recourse to what can be inferred from such data.Educational data mining can make more senses from such data by going beyond just facts and figures in the school settings to home, society, economy, politics, health and others sectors. This is why it is called interdisciplinary educational data miningbecause it can involve the analysis of social networks, educational psychology, cognitivepsychology, psychometrics among others.[3] illustrated the main areas concerned in educational data mining diagrammatically.
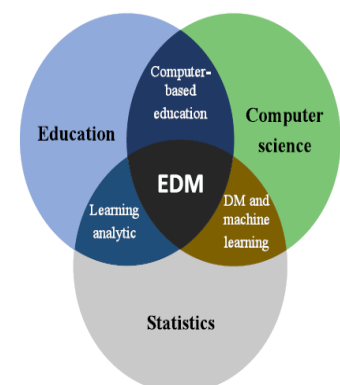

Figure 1: Main Areas Concerned in Educational Data MiningSource: [3]

OBJECTIVES OF DATA MINING

The data mining exercises are carried out with the aim of discovering patterns of relevant and interesting information in large volumes. This is done with the development of four phases [4], these phases are:

1. Filtering data.
2. Selection of variables.
3. Extracting knowledge.
4. Interpretation and evaluation.

From the phases, it is clear that data mining in education can analyse the data generated by any system of learning and focus on diverse aspects, both individual and group and take into account underlying, administrative, demographic and motivational data which in turn contain multiple levels of hierarchy, contexts, levels of granularity and historical data.

METHODS OF EDUCATIONAL DATA MINING

Educational data mining methods stand out from methods of broader data mining going by literature. In other words, methods from the psychometrics literature are often integrated with methods from the machine learning and data mining literatures to achieve goal of educational data mining [5]. For instance, in mining data about how students employ computer assisted learning software (e.g CAI drill and practice), it is important to look at data at different levels at the same time such as login level, reading level, practice level, answer level, session level, and timing level because all these play important roles in the study of educational data.

Technically, there are a wide variety of current methods popular within educational data mining. Thesemethods can fall into the following general categories: prediction, clustering, relationship mining, discovery with models, and distillation of data for human judgment. The first three categories are largely acknowledged to be universal across types of data mining (even though they called with different names in some other literature). The fourth and fifth categories achieve particular prominence within educational data mining.[3]listed the methods as web mining methods. These methods are quite prominent both in mining of web information data and in mining presently different types of educational data. A second perspective on EDM is specified by [1], which analyses and classifies various tasks in EDM as follows:

i. Association rule mining
ii. Causal data mining
iii. Classification
iv. Clustering
v. Correlation mining
vi. Density estimation
vii. Discovery with models
viii. Distillation of data for human judgment
ix. Prediction
x. Regression
xi. Relationship mining
xii. Sequential pattern mining

APPLICATIONS OF EDUCATIONAL DATA MINING IN EDUCATION SECTORS

[6] And [7] highlight very many applications of educational data mining in the education sectors among which are:

i. Analysis and Visualization of Data
ii. Enrolment Management
iii. Infrastructure Management Tuition and fee Management

iv.   Predicting Students' Profiling
v.   Predicting Student Performance
vi.   Predicting Students Career
vii.   Grouping Students
viii.   Planning and scheduling
ix.   Organization of Syllabus
x.   User Modeling
xi.   Detecting Cheating in Online Examination

SOME STATISTICAL SOFTWARE TOOLS FOR EDUCATIONAL DATA MINING

i.   **WEKA (Waikato Environment for Knowledge Analysis):**
The Weka workbench consists of several tools, algorithms and graphics methods that lead to the analysis and predictions. Most of the algorithms are inbuilt in this tool.

ii.   **Intelligence Miner (IBM):**
Intelligence miner provides tight integration with IBB's DB2 relational db system, Scalability of Mining Algorithm. It can be run on Windows, Solaris and Linux.

iii.   **SPSS Clementine (IBM):**
Provides an integrated data mining development environment for end users and developers. It is used for Association Mining, Clustering, Classification, Prediction and visualization tools and run windows and Linux

iv.   **KEEL (Knowledge Extraction Based on Evolutionary Learning):**
KEEL as an application is a set of machine learning software that is designed for providing a resolution to numerous data mining problems. It has a collection of software techniques that are involved in data manipulation and analysis before and after the process as well. It applies soft-computing methods in extracting information about learning and knowledge.

v.   **R (Revolution):**
This is a statistical computing software/ language that is widely used by data miners to perform statistics for learning development solutions. R is an extremely versatile tool that is not only scientifically designed but is also easy to use. So, applying stats and formulas in R are convenient.

vi.   **iDataAnalyzer (Microsoft):**
Provides platform for visual learning environment. This is done through Pre-processor, ESX, Heuristic Agent, Neural Network, Rule Maker and Report generator methods.

vii.   **KNIME (Konstanz Information Miner):**
This platform is a widely used open source for data analytics, reporting, and integration. Traditionally used for pharmaceutical research, this business analysis tool is now widely used for Educational Data Mining.

viii.   **Enterprise Miner (SAS Institute):**
Provides variety of statistical analysis tools through Association Mining, Classification, Regression, Time series analysis, Statistical analysis and Clustering.

ix.   **SIPINA (RiccoRakotomalala Lyon, France):**
Provides an environment for supervised learning algorithms, handle both continues and discrete data.

x.   **ORANGE:**
Orange is a component-based data mining software suite that is suitable for explorative data analysis, visualization, and predictions. It operates perfectly for various exploration techniques and also aids in scoring and filtering data as a part of the post-processing operation.

xi. **Insightful Miner (Insightful Incorporation:**
Provides visual interface, which allows users to wire components together to create self-documenting programs. It runs on Windows, Solaris and Linux.

xii. **Gene Sight (Inc. of EI Segundo, CA):**
Provides the researcher to explore large data sets from multiple experimental groups using advanced normalization, visualization, and statistical decision support tools.

CONCLUSION

Educational Data Mining EDM is far from being reported as concluded it is still very much evolving particularly in developing countries like Nigeria which its applications is very rare. Considering the opportunities offered by the educational data mining to students, teachers, parents, school management and owners, government and philanthropies of education it expedient that expertise in collecting required data, employing appropriate methods, dynamic application in varied spheres of education sectors and adoption of suitable software be enjoined to maximize the opportunities provided by the robust machine learning.

REFERENCES

1. Baker, R. S. & Ryan, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson& B. White (Eds), *Learning analytics: from research to practice*. Berlin, Germany: Springer.
2. Harzing, A. W., &Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*.
3. https://doi.org/10.1007/s11192-015-1798-9.
4. Crist´obal Romero &Sebasti´an Ventura, (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, 2010, pp. 601-618.
5. Virseda Benito, F., & Carrillo, J. (2008). Minería de datos y aplicaciones (p. 8).
6. Ryan S.J.D Baker, (2015). Data Mining for Education.*InternationalEncyclopedia of Education (3rd edition)*. Oxford, UK: Elsevier
7. Prabha, S. L.&Shanavas,A. R.M. (2014). "Educational data mining applications", Operations Research and Applications: An International Journal (ORAJ), Vol. 1, No. 1.
8. Goyal, M. &Vohra, R. (2012)."Applications of Data Mining in Higher Education", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1.