

Research & Survey of Data Mining Techniques In Educational System

Surabhi Juneja, Swapnali Kendre, Parinita Chate

Student, computer engineering Dept, Bharati Vidyapeeths College of Engineering Lavale, Pune, India

Student, computer engineering Dept, Bharati Vidyapeeths College of Engineering Lavale, Pune, India

Associate Professor, computer engineering Dept, Bharati Vidyapeeths College of Engineering Lavale, Pune, India

ABSTRACT: Data mining is the process of finding of hidden information from a huge amount of data. Data mining analyzing the data from different source and convert it into meaningful information. In the world of internet there are several online open source resources that are utilized by various academicians. Unfortunately, all the resources that are available in a scattered manner. Because of this factor it required a proper arrangement of data in meaningful information for this we are going to use data mining process which is popularly known as EDM. Educational data mining is the emerging topic for research community. Educational data mining means to extract the hidden knowledge from large repositories of data with the use of technique and tools. Educational data mining develops new methods to discover knowledge from educational database and used for decision making in educational system. The various techniques of data mining like classification, clustering can be applied to bring out hidden knowledge from the educational data. In this paper, we focus on the educational data mining and classification techniques for a specific sector of Education system.

KEYWORDS: EDM, Database, EDM Components, DM Methods, Education Planning, Data Mining, Profile Management.

I. INTRODUCTION

Educational Data Mining (EDM) is an emerging field exploring data in educational context by applying different Data Mining (DM) techniques/tools. EDM inherits properties from areas like Learning Analytics, Psychometrics, Artificial Intelligence, Information Technology, Machine Learning, Statistics, Database Management System, Computing and Data Mining. It can be considered as interdisciplinary research field which provides intrinsic knowledge of teaching and learning process for effective education. The exponential growth of educational data from heterogeneous sources results an urgent need for research in EDM. This can help to meet the objectives and to determine specific goals of education. EDM objective can be classified in the given way:

Academic Objectives : Person oriented (related to direct participation in teaching and learning process)

E.g.: Student learning, cognitive learning, modelling, behaviour, risk, performance analysis, predicting right enrolment decision etc. both in traditional and digital environment and Faculty modelling- job performance and satisfaction analysis. Department/Institutions oriented (related to particular department/institutions with respect to time, sequence and demand).

E.g.: Redesign new courses according to industry requirements, identify realistic problems to effective research and learning process.

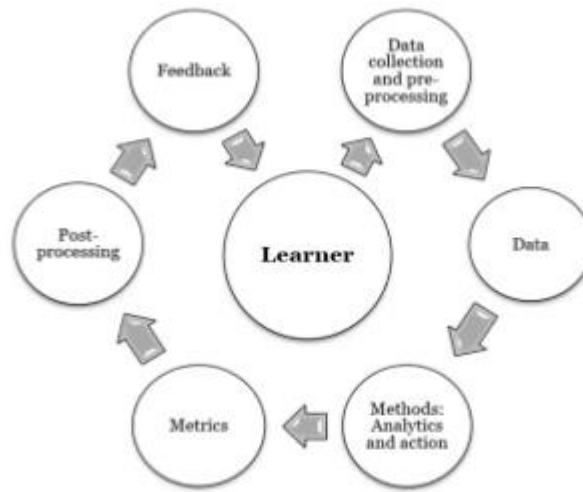
Domain Oriented (related to a particular branch/institutions)

E.g.: Designing Methods-Tools, Techniques, Knowledge Discovery based Decision Support System (KDDs) for specific application, branch and institutions.

II. EDM PROCESS

The process of Educational data mining is an iterative, Knowledge discovery process which consists of Hypothesis formulation, Testing and refinement [4] (see Fig-1). Hypothesis is developed from various educational environments. It creates large volume of data. The main process of EDM starts with validating data (i.e. finding relationship between variables/parameters/data items). This is also known as pre-processing of data. After pre-processing various DM techniques, tools will be employed on processed data and final results/interpretation will be given to different user of education. Further recommendation will be suggested for the refinement of problems/task.

LA/ EDM Process



In LA/ EDM process, data is collected and analyzed, and after post processing, feedback and interventions are made in order to optimize learning (based on [1, 13])

III.EDM MODULES

The main modules of EDM are User and Stake holders of Education, Tools, Techniques and Models of DM, Educational Data, Task and Results that will altogether used to achieve the objectives of EDM. Phases of EDM A. Users and Stakeholders: According to researchers there are four main groups of users and stakeholders involved in accomplishing the specific objectives.

1) Learners (Student/Pupils): The main objective of Learners is to reply to student requisites, improve learning performance, and provide feedback or recommendations to learners. 2) Educators (Teachers/Faculties/Tutors): Their objective is to analyse student behaviour, learning, cognitive and social aspect and reflect on their own teaching methods. Educational Researchers: The main objective of researchers is to develop new tools and techniques for the growth of educational system. Administrators: The main objective is to utilize and enhance available resources (human and material) and their educational offer and so on. B. EDM Methods: One of the essential modules of EDM is the methods of DM, used for different purpose. Romer Ventura and Ryan Baker categorized the methods as- , Prediction Clustering, Relationship mining These methods are useful in mining web data and in mining other forms of educational data. Universally accepted across types of DM. Some of the methods which are acknowledged when validated relationships are applied to make predictions are- Distillation of data for human judgment Discovery with models Knowledge Tracing(KT)

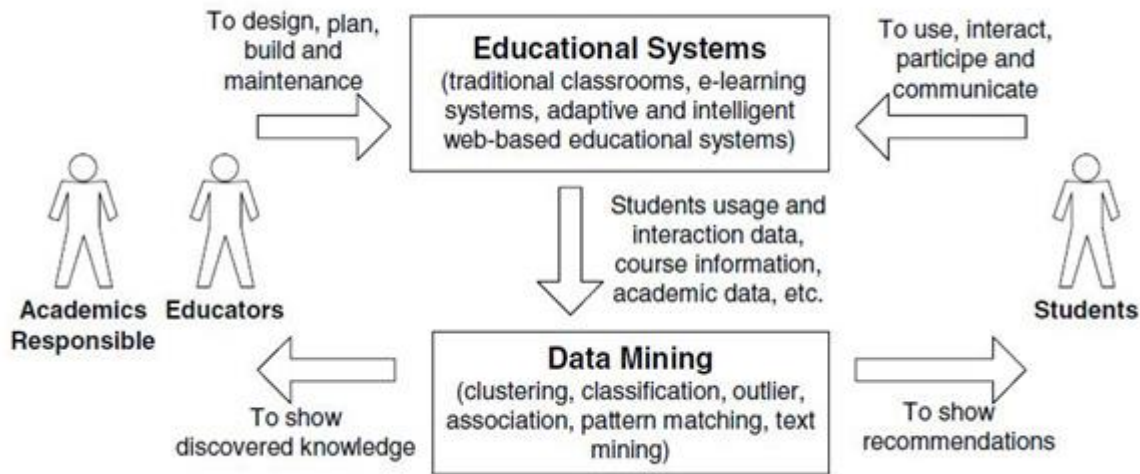
i) Prediction: This technique is used to derive predicted variable (single variable) from predictor variables (combination of variables). Prediction is used analyse student performance and drop out. and for detecting student behavior. It is classified into three types. Classification: used to predict class label from (discrete or continue). Some popular classification methods include logistic regression, support vector machines and decision trees.

Regression: used to predict from continuous variable. Some popular regression methods within educational data mining include linear regression, neural networks. Density Estimation: probability density function is used to predicted variable. Density estimator can be based on variety of kernel functions, including Gaussian function.

ii) Clustering: Clustering is an unsupervised classification process .it is used for grouping objects into classes of similar objects . Data items are partitioned into groups or subsets (clusters) based on their locality and connectivity within N dimensional space. In educational data mining, clustering has been used to group students according to their learning

iii) Relationship mining: Relationship mining is used to determine relationship between variables in a data set and form rules for specific purpose. Relationship mining is classified into four types: Association rule mining: This method is used to identify relationship between attributes in data set, extracting interesting correlations, frequent patterns among data items. for finding students' mistakes often occurring together while solving exercises .Correlation mining: This method is used to find Linear correlations between variables (positive or Negative). Correlation analysis is used to find the most strongly correlation attributes. Sequential pattern mining: This method is used to find intersession patterns such as the presence of a set of items followed by another items in a time-ordered set of sessions or episodes. based on temporal relationship between variables to predict which group a learner belongs to. Wang et. al. proposes a four phase learning portfolio mining approach. Casual data mining: This

method is used to find casual relationship between variables by analysing the covariance of two events or by using information about how one of the events was trigger.



IV. RELATED WORK

In the world of internet there are several online open source resources that are utilized by various academicians. Unfortunately, all the resources that are available in a scared manner. There are various unknown open source resources that are available over the world wide web and can be a proven useful utility for the productivity. Such unknown resources are suffering from the lack of popularization among engineering academic .In today's world ,an engineer can become more productive if the/she is aware of such auspicious information that is provided in a united manner. We are aimed to develop an innovate cloud based digitalized problem that will provide the unified information to the users. This will allow avoid wandering for information over the internet. Eventually this will not only benefit the academicians in a regional way but in a global way. Use of cloud for centralized storage of information will provide the accesstotheinformaontotheend-userinamoreprofoundandsecureway.Thisplaormwillnotonlybe providing just an information about a particular resource but will also various beneficial resources for the end-user. One of the objective of this integrated problem is to provide information in a classified manner. This classification can be done on the basis of stream or subject of engineering.

V. MATHEMATICAL MODEL

Let 'S' be Universal Set that represent the system of soluons for the above described problem.

Then, mathemacally we can say that

$$S = \{s, e, x, y, DD, NDD, Memsh, F, SCase, FCase, CPUcount\}$$

Where,

s : Start State

e : End State

x : Set of Inputs

y: Set of Outputs

DD: Determinisc Data

NDD: Non-Determinisc Data

Memsh: Shared Memory

F: Set of Funcons SCase: Success Case Fcase: Failure Case CPUcount: CPU Core Count

Here,

s = main ()

e = End State

x = {X1, X2, X3}

where,

X1 = {a, b} for User Login Form

Where,

a = User Name

b = Password X2 = {p, q, r, s, t, u} for User Registraon Form

Where,

p = User Name

Q = Password r = Re-type Password s = E-Mail ID t = Mobile Number

$u = \text{Stream of Engineering X3} = \{v\}$ for Search Engine Where, $v = \text{Keyword to be Searched}$
 $y = \{Y1, Y2\}$

Where,

$Y1 = \text{for Recommendaons}$ $Y2 = \{1, m, n\}$ for manual

Where,

1 = Textual Informaon $m = \text{Video Tutorials}$ $n = \text{Expert Advice}$

DD = 1. User Profile Informaon. 2. Recommendaons.

NDD = 1. Keyword to be Searched.

2. Availability of Desired Resource/Informaon.

$F = \{F1, F2, F3, F4, F5, F6, F7, F8\}$

Where,

F1 = Validate Credenals while Login F2 = Smart Recommendaons F3 = Search Result F4 = Log In / Sign In F5 = Log Out / Sign Out F6 = Track Downloads F7 = Noficaons F8 = Fetch Desired Data from Cloud and Display it Scase = 1.

Desired Resource / Informaon is found. 2. Successful Login / Sign In. 3. Successful Log Out / Sign Out.

Fcase = 1. Unavailability of Desired Resource / Informaon. 2. Unable to access Cloud 3. Unable to validate the login Credenals. 4. Unable to support the Client Machine Configuraon.

CPUcount = One Dedicated Server is Used.

II. CONCLUSION AND FUTURE WORK

This paper described about the Educational data mining, goals of educational data mining and phases of educational data mining and existing classification techniques. Various classification techniques can be implemented on the data set but which classification technique will be applied on the data to improve the academic performance of students, it is important. In this paper, we made a comparison analysis on different existing approaches and methods of classification of data sets. We did the comparative analysis on the basis of accuracy percentage on the application of various classification techniques like Naïve Bayesian Classification, Multilayer Perceptron, J48 and ID3 etc. we also analyzed the advantages and shortcomings of each algorithm applied to data set. So we can say that this paper will provide a beneficial glance of existing solution for classification with their advantages and shortcomings.

REFERENCES

- [1] Kumar S. Anupama and N. vijaylaxmi M.2011 Efficiency of Decision trees in predicting Student's Acedemic performance.
- [2] L. Ramanathan, Dhanda S. and D. S. kumar 2013 Predicting Student's Performance using Modified ID3 Algorithm
- [3] Pal A. kumar and Pal S. 2013 Classification Model of Prediction for Placement of Students Amershi, S., Conati, C. and Maclaren, H., (2006) "Using Feature Selection and Unsupervised Clustering to Identify Affective Expressions in Educational Games", in Proc .Of The Intelligent Tutoring Systems Workshop on Motivational and Affective Issues. pp. 21-28.
- [4] Al-shargabi, A.A. and Nusari, A. N (2010), "Discovering Vital Patterns From UST Students Data by Applying Data Mining Techniques", in Proc. Int. Conf. On Computer and Automation Engineering, China: IEEE, 2010, 2,547-551.DOI:10.1109/ICCAE.2010.5451653.
- [5] Badri, M., and El Mourad, T (2011) "Measuring job satisfaction among teachers in Abu Dhabi: design and testing differences" in Proc.Int. Conf. on NIE, 4th Redesigning Pedagogy,Singapore.
- [6] Baker, R.S, Corbett, A.T., Koedinger, K.R (2004) , "Detecting Student Misuse of Intelligent Tutoring Systems" in Proc. Lecture Notes in Computer ScienceVol.3220,531-540.
- [7] Baker, R.S.J.D.,and Yacef, K.(2009), "The state of Educational Data Mining in 2009:A review and future vision" Journal of Educational Data Mining, Vol.1,No. 1,pp.3-17.
- [8] Nelson,B., Nugent, R., Rupp, A. A.(2012), "On Instructional Utility, Statistical Methodology, and the Added Value of ECD: Lessons Learned from the Special Issue", Journal of Educational Data Mining.Vol.4, No.1,pp.224-230.

BIOGRAPHY

Surabhi juneja is a student of computer engineering Department of Bharati vidyapeeth college of engineering lavale pune ,India . Her research interests are Data mining ,security working on a project of engineering applications apps.

Swapnali kendre is a student of computer engineering Department of Bharati vidyapeeth college of engineering lavale pune ,India . Her research interests are Data mining ,security working on a project of engineering applications apps.

Parinita chate is a professor of computer engineering Department of Bharati vidyapeeth college of engineering lavale pune ,India . Her research interests are Data mining ,computer Networks ,operating system she is working as guide on a project of engineering applications apps.