IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 21, Issue 5, Ser. II (Sep - Oct 2019), PP 16-19
www.iosrjournals.org

Student Academic Performance Prediction using Feature Selection based Bagged ID3 approach

Nisha, Er. Prince Verma
CT Group of Institution/CSE, Jalandhar, 144041, India
CT Group of Institution/CSE, Jalandhar, 144041, India

Abstract:*Educational data mining (EDM) is intended to uncover trends and patterns hidden in the thousands of data sets in an educational system and has drawn the attention of academic authorities for being able to bring benefits to educational institutions. This work aims to investigate in more detail the mining of educational data and to study student behaviour.With the rapid development of the Internet and communication technology, online education has drawn more and more attention, online learning platforms, on the other hand, store massive learner behavioral data and educational data. How to effectively analyse and utilize the data to improve the quality of online education has become a key issue urgently needed to be solved in the field of big data in education (BDE), educational data mining (EDM) is exactly an effective and practical method and means of applying BDE. Therefore, EDM is an important academic research hotspot in the field of EDM.*
Keywords: *Educational Data Mining (EDM), Iterative Dichotomiser (ID3)*

Date of Submission: 24-09-2019 Date of acceptance: 12-10-2019

Paper Organization:
Section 1:Thissectioncomprises of introduction and various application areasregarding educational data mining.
Section 2: includes literature review regardingtheexisting work.
Section 3:This section gives conclusion regarding the proposed work.

I. Introduction

Data mining or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information.It works on the principle of retrieving relevant information from data. It is mainly in use of as in Market AnalysisBanks, Insurance companies, Retail store , Hospitals, Customer

II. Retention

A. Applications

1. Educational Data Mining (EDM): EDMis the field of study concerned with mining educational data to find out interesting patterns and knowledge in educational organizationsOff-Line Education (Traditional Class room)In this we need to recollect college students 'conduct, overall performance, curriculum, and so on. That became accrued in study room surroundings.
2. E-Learning and LearningManagement System (LMS): E-Learning gives online guidance. LMS also offers conversation, collaboration, administration and reporting tools.
3. Web Mining (WM) techniques had been applied to students 'information saved via those structures in log documents and databases.
4. Intelligent Tutoring System (ITS) and Adaptive Educational Hypermedia System (AEHS): It adapts teaching to the desires of every unique scholar.
5. Tree Decision Tree-shaped structures that represent sets of decisionsclassifies data using the attributesTree consists of decision nodes and decision leafs. Nodes can have two or more branches which represents the value for the attribute tested. Leafs nodes produces a homogeneous result.These decisions generate rules for the classification of a dataset.Specific decision tree methods include ID3, C45, Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).Phases:A decision tree is created in two phases:

Tree Building Phase

Repeatedly partition the training data until all the examples in each partition belong to one class or the partition is sufficiently small.

Tree Pruning Phase
Remove dependency on statistical noise or variation that may be particular only to the training set

B. ID3 algorithm

In decision tree learning, ID3 (Iterative Dichotomiser3) is an algorithm invented by Ross Quinlan used to generate a decision tree from a dataset. ID3 is the precursor to the C4.5 algorithm, and is typically used in the machine learning and natural language processing domains.ID3 uses information gain to help it decide which attribute goes into a decision node.

Entropy: It Attempts to create the smallest possible decision tree using Information Gain as Attribute Selection Method.Looking for which attribute creates the most homogeneous branches.information gain is based on the decrease in entropy after a dataset is split on an attribute.

Information Gain (IG): Information gain is based on Claude Shannon's work on information theory. InfoGain of an attribute A is used to select the best splitting criterion attribute. The highest InfoGain is selected to build the decision tree.

Info Gain(A) = Info(D)−Info(A)

III. Literature review

This paper[1] uses decision tree algorithm to analyze factors that affect whether bank customers subscribe to fixed deposits or not, and constructs a decision tree model of influencing factors of bank customers subscribe to fixed deposit business. The results show that the three factors that significantly affect customers' subscription for long-time deposits are the number of employees, duration and month, which greatly reduce the scope of customers that banks push to subscribe for long-term deposits, and are conducive to improving the efficiency of banks.

The proposed methodology [2] consists of the phases like preprocessing, attribute selection, classification based on decision tree and performance evaluation. In the data preprocessing phase, the missing values have been removed. The attributes are remodel into a categorized format using the categorization process. Gain ratio, chi square and information gain feature selection methods are tested on preprocessed data. The suitable attributes selected are predicted using classification techniques. In this paper, one of the classification techniques are described and based on ID3, C4.5 and C5.0 is used to predict the teachers' attainment in educational data mining.

The model [3] is based on Variables in 4 dimensions including transaction frequency, types of products or services traded, transaction amount and client age. And using clustering before classification to divide twenty-five types of outlier customer data into four categories and corresponding marketing strategies also are put forward according to different classification of outlier customer data of a company.

The proposed[4] system is a web based application which makes use of the Naive Bayesian mining technique for the extraction of useful information. The experiment is conducted on 700 students' with 19 attributes in Amrita VishwaVidyapeetham, Mysuru. Result proves that Naive Bayesian algorithm provides more accuracy over other methods like Regression, Decision Tree, Neural networks etc., for comparison and prediction. The system aims at increasing the success graph of students using Naive Bayesian and the system which maintains all student admission details, course details, subject details, student marks details, attendance details, etc. It takes student's academic history as input and gives students' upcoming performances on the basis of semester.

To improvise educational data mining, clustering will be used in the[5] paper. As we need to improvise performance as well as unambiguousness of obtained models. We have used 84 under-graduate student data and grouped students according to their final marks they achieved in the course and this we have done by using clustering approach. The result which we get shows that the clarity of specific model is much better than the general model and the unambiguousness of the model is also increase.

This paper [6]present an analysis of the performance of feature selection algorithms on student data set. The obtained results of the different FS algorithms and classifiers will also help the new researchers in finding the best combinations of FS algorithms and classifiers. Selecting relevant features for student prediction model is very sensitive issue for educational stakeholders, as they have to take decisions on the basis of results of prediction models. Furthermore our paper is an attempt of playing a positive role in the improvement of education quality, as well as guides new researchers in making academic intervention.

The purpose[7] of this study is to identify a knowledge to conduct an analysis of student motivation on e-Learning system based on data mining techniques. We use a dataset of the learning system activities in the Program of Distance Learning hosted by the second batch of APTIKOM Consortium. In this study, the association rules and classification techniques are used to identify patterns of knowledge and reorganize the virtual course based on patterns found. The expected result of this research can contribute to a model of Data

Pre-Preparation Process and its steps from Moodle log data as a reference for the researchers. For the future research, we recommend using different datasets from techniques to comprehend divers of result expectations.

This work [8]aims to investigate in more detail the mining of educational data and to suggest a preliminary classification scheme. As a result, we hope to provide an overview of such research area by identifying key topics, types and trends of preliminary research, as well as the maturity of existing contributions. Since the EDM is part of an interdisciplinary area, mobilizing mainly knowledge of statistics, machine learning, pattern recognition, etc., we believe that with the present work one can have a better understanding of the area and its aspects.

This document [9] presents the results of conducting a Systematic Mapping Study (SMS) on the mining of educational data to determine factors that affect school performance in the higher education system. As a result of this analysis, 20 primary studies were obtained, where it is observed how data mining is applied to predict school performance and thus decrease the dropout rates of students.

This paper [10] willanalyze the realization of Web content mining and Web structure mining, their basic algorithm principles and their application areas.

This paper [11] takes the composition of the intelligent teaching system and the realization process of the system as study core, to introduce the process of the design and realization of each component of the intelligent tutoring system. It adopts data mining association rules and decision tree mining algorithm to enhance the intelligence and personalization of the intelligent tutoring system.

The paper [12] introduces the basic concepts of BDE, EDM and online learning platform, and then elaborates on the process of how educational data mining transforms raw data into knowledge. Finally, the key technologies of data mining are classified according to their uses, and gives its application in online education scene. The paper can provide some guidance for the research and application of educational data mining based on online education.

IV. Comparison

Comparison between related work is made for which different techniques were used. For this proposed work decision tree technique and id3 is used.

V. Conclusion

The paper finishes withthedetailed analysis of various techniques and algorithms used for the related work in the field of educational data mining using various techniques such as decision tree technique,id3 etc. Comparison table is used to understand various techniques and approaches discussed in the field of educational data mining.

Refer ence	Author	Technique	Year
[1]	JunfengGuo ; Handan Hou,	decision tree algorithm	2019
[2]	R. Lawrance ; V. Shanmugarajeshwari	ID3, C4.5 and C5.0	2017
[4]	TismyDevasia ;Vinushree T P	Naive Bayesian mining	2016
[6]	Maryam Zaffar ; ManzoorAhmedHashmani ; K. S. Savita	Feature Selection	2017
[9]	Viviana Cristina PárragaVillama	data mining techniques used, such as logistic regression, decision trees, random forests, Naive Bayes	2018

References

[1]. JunfengGuo ; Handan Hou, “Statistical Decision Research of Long-Term Deposit Subscription in Banks Based on Decision Tree”, 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Jan. 2019

[2]. R. Lawrance ; V. Shanmugarajeshwari, “An assay of teachers' attainmentusing decision tree based classification techniques,” 2017 International Conference on Circuit ,Power and Computing Technologies (ICCPCT), April 2017

[3]. Lei Zuo ; JunfengGuo,,Customer Classification of Discrete Data Concerning Customer Assets Based on Data Mining, “2019

[4]. TismyDevasia ;Vinushree T P ; VinayakHegde, “Prediction of students performance using Educational Data Mining”,2016 International Conference on Data Mining and Advanced Computing (SAPIENCE),March 2016.

[5]. AkanshaMishra ;RashiBansal ; Shailendra Narayan Singh, Educational data mining and learning analysis “: 2017 7th International Conference on Cloud Computing, Data Science & Engineering – Confluence, June 2017

[6]. Maryam Zaffar ; Manzoor Ahmed Hashmani ; K. S. Savita , “Performance analysis of feature selection algorithm for Conference on Big Data and Analytics (ICBDA), Nov. 2017

[7]. NurulHidayat ;RetantyoWardoyo ; SN Azhari, “Educational Data Mining (EDM) as a Model for Students' Evaluation in Learning Environment”,2018 Third International Conference on Informatics and Computing (ICIC), Oct. 2018

[8]. Victor Regis LyraBeserra da Silva ; Fábio de Albuquerque Silva ; VanilsonBurégio,,”Characterizing Educational Data Mining”,2019 14th Iberian Conference on Information Systems and Technologies (CISTI),June 2019

[9]. Juan PábloZaldumbideProaño ; Viviana Cristina PárragaVillama,"Systematic Mapping Study of Literature on Educational Data Mining to Determine Factors That Affect School Performance", 2018 International Conference on Information Systems and Computer Science (INCISCOS), Nov. 2018
[10]. Yeqingi, "Research on Technology, Algorithm and Application of Web Mining," 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), July 2017
[11]. YixuanChen ; Yang Zhang," Research on Intelligent Tutoring System Based on Data-Mining Algorithms", 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Jan. 2019
[12]. Wei Zhang ;Shiming Qin, "A brief analysis of the key technologies and applications of educational data mining on online learning platform",2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA),May 2018

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Nisha. "T Student Academic Performance Prediction using Feature Selection based Bagged ID3 approach." IOSR Journal of Computer Engineering (IOSR-JCE) 21.5 (2019): 16-19