

IDENTIFICATION OF SLOW LEARNERS WITH CATEGORIZATION DATA MINING TECHNIQUES

S. Deepa, Dr. R.Priya

¹ M.phil Scholar, Department of School of computing Sciences, Vels Institute of Science, Technology & Advanced studies, Chennai, Tamilnadu, India

² Assistant Professor, Department of School of computing Sciences, Vels Institute of Science, Technology & Advanced studies, Chennai, Tamilnadu, India

deepaaranathan@gmail.com, priyaa.research@gmail.com

Abstract— The applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining (EDM). The biggest challenge is to improve the quality of the educational processes so as to enhance student's performance. Thus, it is crucial to set new strategies and plans for a better management of the current processes. It is concerned with developing methods for exploring the unique types of data that come from educational environments. This research work detail with of slow learner students in school education. There are so many affecting factors like economy, environment, social media, psychological issues, child hood marriage, irregular attendance and eve-teasing of girls. The goal is to have a better understanding of how students learn and identify the settings in which, they learn to improve educational outcomes and to gain insights into and explain educational phenomena. Many studies have been attempted to put together a process that will identify student at risk of slow learners using various research methodologies. The purpose of this research work is to investigate and correlate the school slow learners with the use of data mining techniques of existing data sources with J48 algorithm. Experiment attempt to improve their accuracy for predicting which student might not attend the school. Using all the available attributes, selection of the best attributes is made and finally rebalancing of data is done using cost sensitive classification. The slow learner students result and attendance have been compared and the models with the best results are shown.

Keywords- Data Mining, Educational DM, Prediction, Classification, Clustering

I. INTRODUCTION

Educational data has become a vital resource in this modern era, contributing much to the welfare of the society. Educational schools are becoming more competitive because of the number of school which is growing rapidly. To stay afloat, the educational schools are focusing more on improving various aspects and one important factor among them is quality learning. For providing quality education and to face new challenges, the school needs to know about their potentials which are explicitly seen and which are hidden. How to improve the overall education system? There are lots of factors which affect the education system like the student, management, administrators, infrastructures, teachers,

teaching methodology, basis facilities for boys and girls (like separate toilets for M/F), transportation facilities etc. So with the help of educational data mining techniques, detailed analyses are performed on these factors and find out those factors which affect the education of the student and hence students underperformed. The truths behind today's educational school are a substantial amount of knowledge is hidden. To be competitive, the school should identify their own potentials hidden and implement a technique to bring it out.

Prediction Process Model (PPM): The output from analysis process will help the school student to predict academic trends and patterns by categorizing the students into good, satisfactory, or poor group. The model is mainly focused on exploring various indicators that have an effect on the academic performance of the students. The extracted information that describes student performance can be stored as intelligent knowledge for decision making to improve the quality of education in school. The knowledge stored is used for predicting the slow learner student's performance in advance.

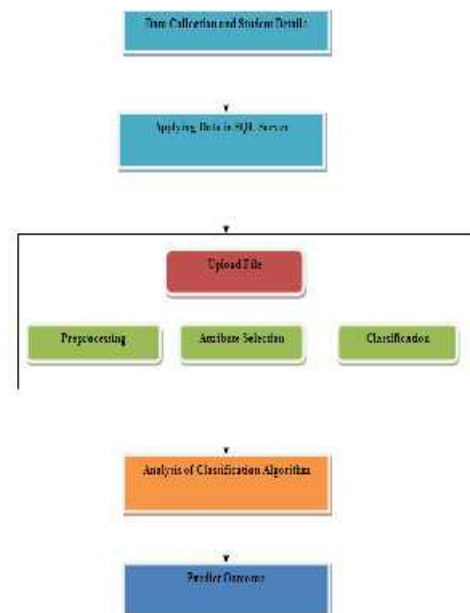


Fig 1: The Workflow Diagram

Data Mining Using Naïve Bayes And J48 Algorithm:

Naïve Bayes Algorithm Process: It was possible to identify the main determining attributes/variables and their values for the failure or success of students in a specific school; number of students in a class, number of subject given in a particular class, to collect the slow learner examination result of a student were the main determining attributes obtained from this research result.

J48 Algorithm Process: In other algorithms the classification is performed recursively till every single leaf is pure, that is the classification of the data should be as perfect as possible. This algorithm it generates the rules from which particular identity of that data is generated. The objective is progressively generalization of a decision tree until it gains equilibrium of flexibility and accuracy.

Working Process: The dynamic interface can also use the collect and arrange school student details. The classification algorithms Naive Bayes, decision tree (J48) they were used to make predictions. The dynamic interface can also use the constructed models that mean the application worked properly in each considered case. After the tree is fully constructed, this process of algorithm.

Experimental Setup: In this process, a questionnaire form is used to collect the real data from the students that describe the relationship between learning behaviour and their academic performance. The variables for judging the learning and academic behaviour of students used in the questionnaire are student demographic details, School details, Attendance, percentage and Final grade in last examination. These data's are thereby recorded in excel sheets for analysis. Data sets about 300 students were collected. Among the dataset around 200 are used as training dataset and 50 datasets as test data to design student model.

II. LITERATURE REVIEW

2.1 Mining Educational Data to Analyze Student Performance

The classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the student's previous database, to predict the performance at the end of the semester. This study will help to the students and the teachers to improve the division of the student.

2.2 A CHAID Based Performance Prediction Model in Educational Data Mining

The features like medium of instruction, marks obtained in secondary education, location of school, living area and type of secondary education were the strongest indicators for the student performance in higher secondary education. This

CHAID prediction model of student performance was constructed with seven class predictor variable, whereas the earlier models in reviews were constructed with limited class predictor variables. A hybrid model, which utilizes the full impact of variable selection and its consequences, is being worked out and it is expected to present a better picture of the real scenario.

2.3 A Study of Application of Data Mining and Analytics in Education Domain

Data mining techniques like outlier analysis, naïve Bayesian algorithm should be used to research into are-as like finding students with similar interests in order to pursue research projects in that sector and to identify groups of students in need of extra guidance in particular areas to schedule extra coaching sessions. Analysis of students should be carried out on a weekly basis by continuously assessing them through assignments and research projects. A recently introduced concept of academic analytics uses the data mining algorithms on the educational data of students and gives certain insights about the expected performances of the students, expected retention rate of students and percentage of resources properly utilized.

2.4 Using Data Mining To Predict Secondary School Student Performance

The present work intends to approach student achievement in secondary education using BI/DM techniques. Recent real-world data (e.g. student grades, demographic, social and school related features) was collected by using school reports and questionnaires. On the other hand, the fields of Business Intelligence (BI)/Data Mining (DM), which aim at extracting high-level knowledge from raw data, offer interesting automated tools that can aid the education domain. The present work intends to approach student achievement in secondary education using BI/DM techniques.

111. DATA COLLECTION

3.1 Data Set

To collect the slow learner information in school. The scope of educational data mining includes areas that directly impact students. Data mining should be applicable to any kind of information repository.

	A	B	C	D	E	F	G	H	I	J
	Year	Index	D0	Rate	5-year	10-year	20-year	30-year	Monthly	Quarterly
10/1/2000	2000	2000	0.00	0	0	0	0	0	0	0
11/1/2000	2000	2000	0.00	0	0	0	0	0	0	0
12/1/2000	2000	2000	0.00	0	0	0	0	0	0	0
1/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
2/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
3/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
4/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
5/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
6/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
7/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
8/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
9/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
10/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
11/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
12/1/2001	2001	2001	0.00	0	0	0	0	0	0	0
1/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
2/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
3/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
4/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
5/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
6/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
7/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
8/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
9/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
10/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
11/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
12/1/2002	2002	2002	0.00	0	0	0	0	0	0	0
1/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
2/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
3/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
4/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
5/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
6/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
7/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
8/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
9/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
10/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
11/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
12/1/2003	2003	2003	0.00	0	0	0	0	0	0	0
1/1/2004	2004	2004	0.00	0	0	0	0	0	0	0
2/1/2004	2004	2004	0.00	0	0	0	0	0	0	0
3/1/2004	2004	2004	0.00	0	0	0	0	0	0	0
4/1/2004	2004	2004	0.00	0	0	0	0	0	0	0
5/1/2004	2004	2004	0.00	0	0	0	0	0	0	0
6/1/2004	2004	2004	0.00	0	0	0	0	0	0	0

Educational data mining (EDM) is an emerging discipline which focuses applying data mining tools and techniques to educationally related data. An educational data mining is a broader term that focuses on nearly any type of data in educational institutional, while academic analytics is specific to data related institutional effectiveness and student retention issues. The scope of educational data mining includes areas that directly impact students. In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: Classification, categorization, estimation, and visualization. Classification identifies associations and clusters, and separates subjects under study.

Column Name	Data Type	Allow Nulls
name	varchar(50)	☐
gender	varchar(50)	☐
dob	varchar(50)	☐
class	varchar(50)	☐
subjects	varchar(50)	☐
health_related_problems	varchar(50)	☐
father_occupation	varchar(50)	☐
affected_by_any_fma...	varchar(50)	☐
income	varchar(50)	☐

G015\SQLXPRES... dbo.data_set		Summary
Column Name	Data Type	Allow Nulls
name	nvarchar(50)	☐
gender	nvarchar(50)	☐
dob	nvarchar(50)	☐
class	nvarchar(50)	☐
subjects	nvarchar(50)	☐
health_related_problems	nvarchar(50)	☐
father_occupation	nvarchar(50)	☐
affected_by_any_fma...	nvarchar(50)	☐
income	nvarchar(50)	☐

UNIT 54 - 4 - 10th grade - 2010									
id	name	age	sex	height	weight	eye color	hair color	skin color	birth
1000000001	John	10	M	140	35	Brown	Black	White	2000
1000000002	Jane	10	F	135	30	Blue	Blond	White	2001
1000000003	Mike	10	M	145	40	Green	Black	White	2002
1000000004	Sarah	10	F	138	32	Blue	Blond	White	2003
1000000005	David	10	M	142	38	Brown	Black	White	2004
1000000006	Emily	10	F	136	31	Blue	Blond	White	2005
1000000007	Chris	10	M	141	39	Brown	Black	White	2006
1000000008	Alex	10	M	143	41	Brown	Black	White	2007
1000000009	Anna	10	F	137	33	Blue	Blond	White	2008
1000000010	Tom	10	M	144	42	Brown	Black	White	2009
1000000011	Lisa	10	F	139	34	Blue	Blond	White	2010
1000000012	Ben	10	M	146	43	Brown	Black	White	2011
1000000013	Mia	10	F	140	36	Blue	Blond	White	2012
1000000014	Ethan	10	M	147	44	Brown	Black	White	2013
1000000015	Sophia	10	F	141	37	Blue	Blond	White	2014
1000000016	Noah	10	M	148	45	Brown	Black	White	2015
1000000017	Isabella	10	F	142	38	Blue	Blond	White	2016
1000000018	Lucas	10	M	149	46	Brown	Black	White	2017
1000000019	Olivia	10	F	143	39	Blue	Blond	White	2018
1000000020	Max	10	M	150	47	Brown	Black	White	2019
1000000021	Charlotte	10	F	144	40	Blue	Blond	White	2020
1000000022	Leo	10	M	151	48	Brown	Black	White	2021
1000000023	Aria	10	F	145	41	Blue	Blond	White	2022
1000000024	Jack	10	M	152	49	Brown	Black	White	2023
1000000025	Grace	10	F	146	42	Blue	Blond	White	2024
1000000026	Henry	10	M	153	50	Brown	Black	White	2025
1000000027	Victoria	10	F	147	43	Blue	Blond	White	2026
1000000028	Sebastian	10	M	154	51	Brown	Black	White	2027
1000000029	Madison	10	F	148	44	Blue	Blond	White	2028
1000000030	Christopher	10	M	155	52	Brown	Black	White	2029
1000000031	Chloe	10	F	149	45	Blue	Blond	White	2030
1000000032	Matthew	10	M	156	53	Brown	Black	White	2031
1000000033	Amelia	10	F	150	46	Blue	Blond	White	2032
1000000034	Isaac	10	M	157	54	Brown	Black	White	2033
1000000035	Evelyn	10	F	151	47	Blue	Blond	White	2034
1000000036	Samuel	10	M	158	55	Brown	Black	White	2035
1000000037	Abigail	10	F	152	48	Blue	Blond	White	2036
1000000038	Benjamin	10	M	159	56	Brown	Black	White	2037
1000000039	Madeline	10	F	153	49	Blue	Blond	White	2038
1000000040	Joseph	10	M	160	57	Brown	Black	White	2039
1000000041	Elizabeth	10	F	154	50	Blue	Blond	White	2040
1000000042	David	10	M	161	58	Brown	Black	White	2041
1000000043	Sophia	10	F	155	51	Blue	Blond	White	2042

The field of Data Mining is concerned with finding new patterns in large amounts of data. Widely used in Business, it has scarce applications to Education. The amount of data stored in educational databases is increasing rapidly. In order to get required benefits from such large data and to find hidden relationships between variables using different data mining techniques developed and used. This research work addresses the capabilities of data mining and its applications in education.

Education data mining main focus is to analyze the education system. The model focuses on analyzing the prediction accuracy of the student's performance .The dataset that comprises of all academic and personal factors of the students. Thereby improve their standards and reputations. As a result the quality of education can be improved. The results of the data mining algorithms for the categorization of the students based on the attributes selected reveals that the prediction rates are not uniform among the algorithms. The range of calculation varies from (80-98%).Thereby by comparative analysis of category algorithms (such as Naïve bayes, J48) using SQL(Structure Query Language) tool, it is proven that the attributes chosen from the original dataset have high influence using J48 with an accuracy of 97% under analysis and used for predicting test data set for future outcome as best, good, average or poor.

The work can be further extended out by designing the student analyzing, records of students extra-curricular skills and provide a suggestions on communication and technical

skill development by which students can be build in professional aspect of talents.

V1. REFERENCES

- [1] Han Jiawei, Micheline Kamber, "Data Mining: Concepts and Technique". Morgan Kaufmann Publishers, 2000.
- [2] Weiss S. & Indurkha N, "Predictive Data Mining: A Practical guide", Morgan Kauf-. Mann, 1998.
- [3] Technology Forecast: 1997 (1997), Price Waterhouse World Technology Center, Menlo Park, CA.
- [4] William J. Frawley, Gregory Piatetsky-Shapiro, and Christopher J. Matheus, "Knowledge Discovery in Databases: An Overview", AI Magazine Volume 13 Number 3 (1992) (© AAAI)
- [5] BK Bhardwaj, S. Pal, " Mining Educational Data to Analyze Students Performance", (IJACSA), Vol. 2, No. 6, 2011
- [6] M Bala, Dr. DB Ojha, "Study of applications of Data Mining Techniques in Education", (IJRST) 2012, Vol. No. 1, Issue No. IV, Jan-Mar ISSN: 2249-0604.
- [7] Ying Zhang, Samia Oussena, Tony Clark, Hyeonsook Kim, "Use Data Mining to improve student retention in higher education – A CASE STUDY"
- [8] P. Cortez, and A. Silva (2008), "Using Data Mining To Predict Secondary School Student Performance", In EUROSIS, A. Brito and J. Teixeira (Eds.), pp 5-12.
- [9] Galit.et.al, "Examining online learning processes based on log files analysis: a case study". Research, Reflection, and Innovations in Integrating ICT in Education 2007.
- [10] M. Ramaswami and R. Bhaskaran (2010), "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, pp 10-18.
- [11] Kamal Bunkar, Rajesh Bunkar, Umesh Kumar, Singh Bhupendra Pandya, —Data Mining: Prediction for Performance Improvement of Graduate Students using Classification, Ninth International Conference on Wireless and Optical Communications Networks (WOCN), IEEE conference publication, pp 1-5, Sept 2012
- [12] Carlos Márquez Vera, Cristóbal Romero Morales and Sebastián Ventura Soto,—Predicting of school failure and dropout by using data mining techniques, The, Vol. 8, No. 1, pp 7-14, Feb 2013.
- [13] Suhas G. Kulkarni, Ganesh C. Rampure, Bhagwat Yadav, —Understanding Educational Data Mining (EDM), International Journal of Electronics and Computer Science Engineering, 2013.
- [14] Parneet Kaura, Manpreet Singhb ,Gurpreet Singh Josanc "Classification and Prediction based DataMining Algorithms to Predict Slow Learners in Education Sector" Science Direct Procedia Computer Science 57 (2015) 500 –508 2015 (ICRTC- 2015).
- [15] Renza Campagni, Donatella Merlini, Renzo Sprugnoli, Maria Cecilia Verri, "DataMining Models for Student Careers", Science Direct -Expert Systems with Applications 42 (2015)5508–5521.
- [16] Nurbiha A Shukora , Zaidatun Tasira, Henny Vander Meijden, "An Examination of Online Learning Effectiveness using DataMining", Science Direct - Procedia - Social and Behavioral Sciences 172(2015) 555 – 562.
- [17] Harwatia, Ardita Permata Alfiania, Febriana Ayu Wulandari," Mapping Student's Performance Based on Data Mining Approach", Science Direct Agriculture and Agricultural Science Procedia3(2015) 173 – 177
- [18] V.Ramesh Assistant Professor Department of CSA, SCSVMV University Kanchipuram India "Predicting Student Performance: A Statistical and Data Mining Approach", International Journal of Computer Applications (0975 – 8887) Volume 63– No.8, February 2013 35.
- [19] Krpan, Slavomir Stankov, "Educational Data Mining for Grouping Students in E-learning System", Proceedings of the ITI 2012 34th Int. Conf. on Information Technology Interfaces, June 25 – 28, 2012, Cavtat, Croatia
- [20] Dursun Delen, "A Comparative Analysis of Machine Learning Techniques for Student Retention Management", Science Direct - Decision Support Systems 49 (2010) 498–506.

