

Predicting Eligible Educator Category for Disability Student Welfare using Decision Tree Method

M.Balamurugan, PhD
Associate Professor
School Of Computer Science, Engineering and
Applications,
Baharathidasan University,
Trichy, Tamil Nadu, India.

K.Viji
Research Scholar,
School Of Computer Science, Engineering and
Applications,
Baharathidasan University,
Trichy, Tamil Nadu, India

ABSTRACT

Schools private, public and aided enroll thousands of students into various standards. At the time of admission schools collect the information from the students and store into the computer. Understanding the use of data in student point of view is very important. This work is proposed to analyze the student enrolment data and classify the nature of disability students for admissibility of special welfare scheme using data mining decision tree technique's ID3 algorithm. Decision tree is a method that helps to make good choice, particularly decisions that involve values and risk highly. Decision trees use an explicit methodology to compare challenging alternatives and allocate values to those replacements by combining uncertainties, costs into particular numerical values. ID3 is a popular and most used decision tree algorithm popular for the intrinsic worth of high classifying speed, easy and strong understanding ability and easy creation. Using this method, create the suggestion to the disability student for selecting suitable educator for their studies.

Keywords

Data mining, EDM, Classification, Decision Tree, Student Enrolment, ID3, Weka Tool, Disability Student Welfare.

1. INTRODUCTION

Educational Data Mining is an evolving interdisciplinary research area that deals with the development of methods to explore data creating in an educational context. EDM uses well defined computational methodologies to scrutinize educational data in order to learning educational demands. The data mining technique to explore the educational database is very expected to be of great benefit to the educational institutions. Educational Data Mining (EDM) is a well growing discipline, concerned with data from academic field to develop various methods and to identify unique Education. It is one of the primary requirements for its development. The big advantage of India is having a great population of young students presents a huge opportunity to the players in the education sector at the same time scope to the government for development of this sector consequently [1]. Raising the growth of education sector enrolment also increased and it contains large amount of data. That data can be stored into data warehouse. Data warehouse could be considered as storehouse of data. Data could be collected from different various resources, thus allowing data analysis of different kinds. These different varieties of data would be transformed and integrated to make it similar, and then result making using this data is very easy. Based on the kind of data to handle

with, different data mining algorithms and tools are being used for effective data mining [5]. Patterns will help to explore student's academic information. The process of data mining consists of two stages:

1.1 Exploration Stage

This stage consists of pre- processing data, i.e., before put on data mining algorithms on the data, the school enrolment data sets must be assembled from all disparate sources. In other words the data must be extracted from all sources such that the disparity is eliminated. In this stage the data is cleansed and transformed so that the noise and missing values are dealt with.

1.2 Data mining Stage

This stage takes place after execution study. In this stage, classification decision tree ID3 data mining algorithm will mine the student enrolment data and show recommendation to the disability students for selecting the eligibility educator category for welfare scheme [2].

The following are techniques and tool used for data mining process.

1.2.1 Decision Tree

Decision tree is very easy to understandable by domain experts. A decision tree consists of nodes that exactly have one incoming edge, the root node that has no incoming edges. A node with outgoing edges is called an internal node, the other nodes are called leaves or terminal nodes or decision nodes. ID3 is one of the decision tree algorithms [6].

1.2.2 ID3 Algorithm

ID3 select the splitting attribute using the information gain value. It accepts only categorical attributes but continuous attribute can be handled by converting or partitioning them. In this algorithm the attribute with the maximum information gain value is selected as the root node in building the decision tree [7].

1.2.3 Weka Tool

Weka is a data mining tool to analyze the data. We use this tool to analyze student enrolment data from different state and using decision tree algorithm.

2. LITRATURE REVIEW

In this section describes the source of the problem statement and inspiration of technique and tool used in this paper.

In Dec 2014, Irfan Ajmal Khan and Jin Tak Choi [8] developed the scholarship prediction system using data mining technique to help the institution to select the eligible

student for scholarship based on their performance in the first semester and pin point the student weak area and give better training and guidance to win the opportunity next time.

In Jun 2013, Rupali Bhardwaj, Sonia Vatta [9] have developed a system for describe the importance of ID3 algorithm and implement the ID3 algorithm using training data set.

In Feb 2012, Mohammed M. Abu Tair, Alaa M. El-Halees[5] have a case study about how the mined educational data to improve the performance of the student using different data mining technique.

In Aug 2014, Devikala.D and Kamalraj.N[3] have developed a system for analysis information about academic failure and student dropouts using data mining technique.

In Sept 2014, Shrey Bavisi^{*}, Jash Mehta [^] and Lynette Lopes[^][2] have described about make comparative study of data mining algorithm and discuss the definition of data mining, process of data mining technique. Then describe the two stages of data mining, study of data mining algorithms, advantage, disadvantage of those algorithms.

3. PROBLEM STATEMENT

Decision trees are a form of multiple variable or multiple effect analyses and this analyses allow us to predict, explain, describe, or classify an outcome. ID3 is a classical decision tree algorithm play a vital role in data mining. This work describes the process of ID3 algorithm in education sector to analyze the education data because education is the strength of all growing country. Approximately 70 million disability people live in India. Education is very significant to live normal life for disability people. This paper also talk over the student enrolment information in various state and classify the nature of disability student data then give the tree based result about the data and get knowledge about disability student in schools. Then predict the educator category of school disability student possibility to get welfare. This process can be used to increase the enrolment for nature of disability student, increase teacher for special care for those pupils and easy understanding about the performance of data mining decision tree ID3 algorithm.

4. METHODOLOGY

This paper has used the most common, open-source data mining tool Weka 3.6.9 for this analysis. It contains collection of machine learning algorithms for data mining tasks. Before mining pre-process the school enrolment data through Weka pre-processing technique for reduce the noise and irrelevant or inconsistent data. Give the pre-processed student enrolment input parameter to the Weka classifier and use decision tree ID3 prediction algorithm for result. The ID3 algorithm has two methods entropy and information gain involves processing the input parameters and helps to construct a tree format result.

Table 1 summarizes the parameters of student enrolment. Eligibility is the prediction parameter; the process will be done based on it. Percentage (%) of disability value and eligibility for welfare scheme are the key parameter for construct of the tree and other parameters are help to take decision about disability student welfare scheme such as year, state, school category, student name, gender, nature of disability etc.

Table 2 summarizes steps for the ID3 (Iterative Dichotomiser 3) algorithm to find the decision tree. The mathematical algorithm ID3 can be used for generate the decision tree for given dataset. In ID3 each step makes a decision and it deals symbolic value effectively. This algorithm uses greedy search for select the best attributes and never look back to reconsider the earlier choices. In this method separate the attributes and the possible value of the attributes and it generate the decision tree based on the prediction attribute. Entropy and Gain value can be used to find the root node. Entropy is used for measure the impurity of the attribute. Gain value used to minimize the decision tree for selecting optimal attribute. Formula to calculate Entropy and Gain value, Entropy

$$H(S) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (1)$$

S - The current (data) set for which entropy is being calculated (changes every iteration of the ID3 algorithm), **X** - Set of classes in **S**, **p(x)** - The proportion of the number of elements in class **x** to the number of elements in set **S**. Gain

$$IG(A) = H(S) - \sum_{t \in T} p(t)H(t). \quad (2)$$

H(S) - Entropy of set **S**, **T** - The subsets created from splitting set **S** by attribute **A**, **p(t)** - The proportion of the number of elements in **t** to the number of elements in set **S**, **H(t)** - Entropy of subset. These steps repeated to all attributes to achieve the task. The Weka tree visualizer used to generate the decision tree of student enrolment data for predicting the educator category for disability student welfare.

The algorithms can either be applied directly to a dataset or called from programmable code. Weka encompasses tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Table 1. Student Enrolment Data

S.No	Parameter	Possible Value
1	Year	2010-2012
2	State	Tamil Nadu, Andhra, Kerala, Karnataka, Goa
3	School Category	Private/Public
4	Student name	String value(like Anitha, Mala)
5	Standard	Numeric value(1-10)
6	Gender	Male/Female
7	Nature of disability	Blind, Low Vision, Hearing, Speech, Mentally Retarded, Learning, Cerebral Palsy, Autism, Multiple
8	% of disability	Any numeric value(1-100)
9	Eligibility	Yes/No

Table 2. Steps for ID3 Algorithm

Steps	Process
Step1	To find the entropy value of each attribute for finding root node.
Step2	If entropy >0 , it have a branch Else The attribute doesn't have any branch
Step3	Find gain value for each attribute based on their entropy value. Maximum gain value attribute treated as a root node.
Step4	Find the possible value of root node attribute.
Step5	Find entropy and gain value of each possible value for further movement.
Step6	Continue the step1-step5 iteratively for find the next root nodes.
Step7	Get the tree structured result

This analysis has been performed on a HP Windows system with Intel® Core™ i3 CPU, 2.40 GHz Processor and 4.00 GB RAM.

5. RESULT AND DISCUSSION

The results are obtained using Weka 3.6.9 tool. This paper takes input of student enrolment data from different states that are labeled as charts given below. The figures show the inputs of this process that are year of the enrolment, state of the student, school category either private or public, which standard students studied 1 to 10, gender of the student, student nature of disability, value of disability is the important factor to analyses the data and find educator category of student eligibility for welfare scheme respectively.

5.1 Input Data

The Input Attributes are shown in the form of charts and extracted from the Online Spread Sheet used for clear understanding before Weka 3.6.9 data mining process.

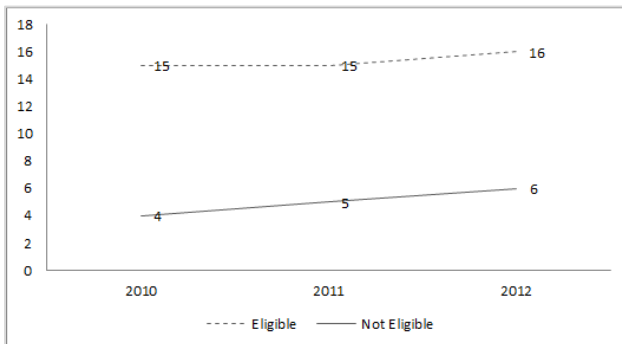


Fig 1: Years

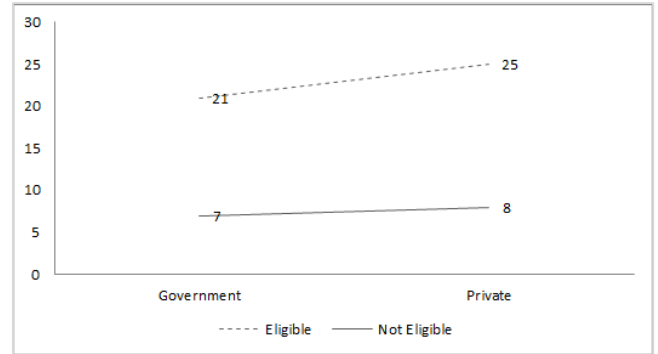


Fig 2: School Category

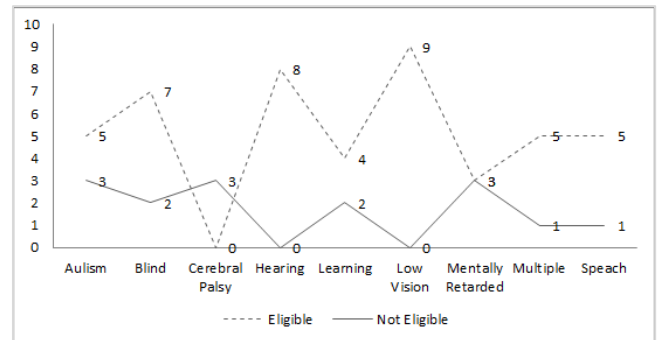


Fig 3: Nature of Disability

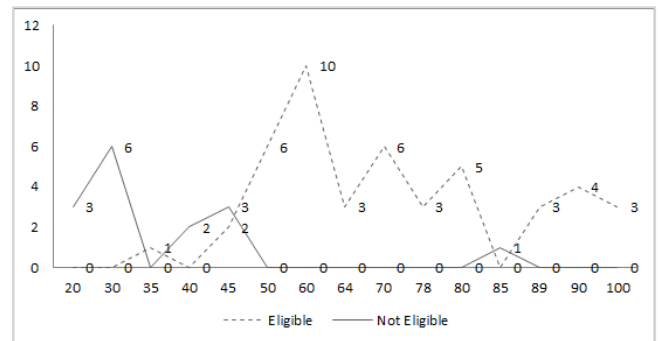


Fig 4: Value of Disability

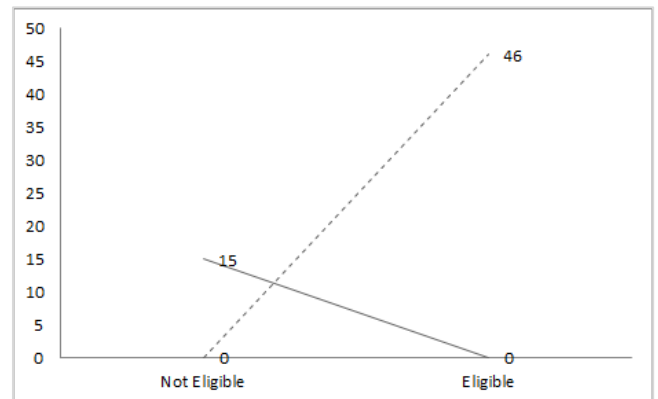


Fig 5: Eligibility for Welfare Scheme

5.2 Output

The output Fig 6 shows the result of prediction for disability student welfare scheme by ID3 decision tree algorithm using Weka classifier. It process the input data and consider the attributes “value of disability” and “eligibility for welfare scheme” as the key for decision making based on the entropy and information gain value of ID3 algorithm. Then produce the result in the form of tree for easy decision making. Finally the results show private sector student having maximum possibility get welfare scheme based on the input.

The Threshold Curve feature generates the points showing prediction tradeoffs that can be obtained by erratic the threshold value among the classes. The resulting knowledge can be used to visualize precision/recall tradeoff or for ROC curve analysis (False Positive Rate vs True Positive Rate). The curve is shown in Fig 7.

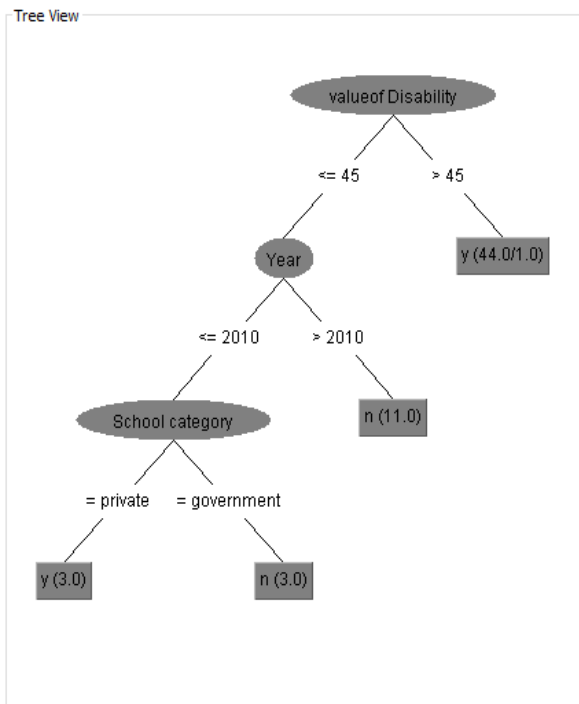


Fig 6: Prediction Results of Student Enrolment Data using Weka 3.6.9

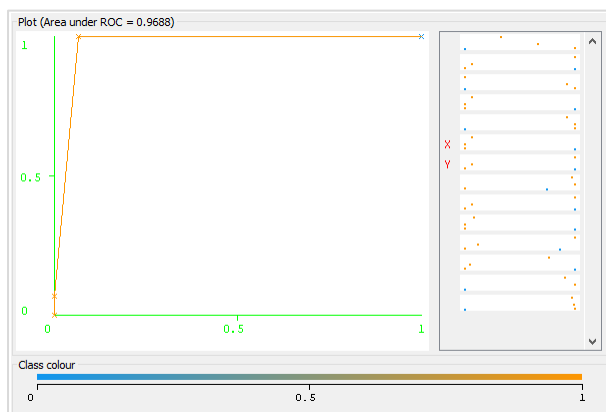


Fig 7: Weka Classifier Visualize: TherosoldCurve (Class Value y)

Table 3. Detailed Accuracy By Class using Weka

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.067	0.979	1	0.989	0.969	y
	0.933	0	1	0.933	0.966	0.969	n
Weighted Avg.	0.984	0.05	0.984	0.984	0.983	0.969	

The accuracy rate can be derived by running ID3 algorithm for loaded input data in Weka 3.6.9.

6. CONCLUSION

This system helps in predicting educator category of school students having maximum possibility to get special disability student welfare scheme. This prediction provides the recommendation to the disability students for selecting the proper school for learning and it will give the knowledge to the academicians and instructors. This prediction makes the competition between the education sectors to increase the disability student enrolment. Increasing the literacy rate for disability people obviously they get better awareness, normal life, more confidence and fine growth. And also the result shows the ID3 decision tree algorithm produce optimal solution to predict the education data and give the better performance for inadequate size of dataset. This system predicts eligible sector for welfare only five states in India. In future we can develop system for all states of India or all countries in world and enhance the performance of ID3 algorithm for handling unlimited size of data based on the more effective and reasonable rule.

7. REFERENCES

- [1] Education Sector in India, India Brand Equity Foundation, www.ibef.org.
- [2] International Journal of Current Engineering and Technology E-ISSN 2277 – 4106, P-ISSN 2347 - 5161 ©2014 INPRESSCO®, “ Comparative Study of Different Data Mining Algorithms” Shrey Bavisia*, Jash Mehta A and Lynette LopesA AComputer Department, DJSCOE, Vile Parle (W), Mumbai – 400056, India.Accepted 02 Sept 2014, Available online 01 Oct 2014, Vol.4, No.5 (Oct 2014)
- [3] International Journal of Computer Trends and Technology– volume 14 number 3 – Aug 2014 -ISSN: 2231-803 “Data Mining Approaches on Detection of Students’ Academic Failure and Dropout: A Brief Survey” -Devikala.D M.phil1 and Kamalraj.N MCA, M.phil2 ,Department of Computer science, Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore, Head of the Department (B.Sc .CT), Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore.
- [4] International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology ISSN 2320-088X ijcsmc, vol. 2, issue. 4, april 2013, pg.374 – 383 research article © 2013, ijcsmc all rights reserved 374 “role of data mining in education sector” Dr. Mohd Maqsood Ali1 1Asst. Professor and Head of Marketing Department, Jazan Community College, Jazan University, Jazan Kingdom of Saudi Arabia 1 maqsoodphd@gmail.com

- [5] Volume 2 No. 2, February 2012 ISSN 2223-4985 International Journal of Information and Communication Technology Research ©2012 ICT Journal. “Mining Educational Data to Improve Students’ Performance: A Case Study” Mohammed M. Abu Tair, Alaa M. El-Halees Faculty of Information Technology Islamic University of Gaza Gaza, Palestine.
- [6] International Journal of Computer Science, Engineering and Information Technology, vol.2, no.2, april 2012.doi : 10.5121/ijcseit.2012.2206 55 “using data mining techniques for diagnosis and prognosis of cancer disease “Shweta Kharya,Sr.Assistant Professor,Bhilai Institute of Technology,Durg-491 001, hhattisgarh, India,Shweta.bitdurg@gmail.com.
- [7] Prediction using Classification Technique for the Students’ Enrollment Process in Higher Educational Institutions-International Journal of Computer Applications (0975 – 8887) Volume 84 – No 14, December 2013.
- [8] An Application of Educational Data Mining (EDM) Technique for Scholarship Prediction, Irfan Ajmal Khan and Jin Tak Choi, International Journal of Software Engineering and Its Applications Vol. 8, No. 12 (2014), pp. 31-42.
- [9] Implementation of ID3 Algorithm, Rupali Bhardwaj , Sonia Vatta CSE, Bahra Uuniversity India, Volume 3, Issue 6, June 2013 ,ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering, Research Paper , Available online at: www.ijarcsse.com
- [10] Vijaykumar S., Saravanakumar S.G., M. Balamurugan. Unique sense: smart computing prototype. Procedia Computer Science. Volume 50, 2015, Pages 223–228. doi:10.1016/j.procs.2015.04.056.