

Educational Mining in Igher Studies Prediction through Decision Tree Algorithm:

P.Anitha Vairamany, Dr.K.Subramaniyan

P.hD Research Scholar, Bharathiyar University,

Professor, Department of Computer Govt arts and science college Pudhukottai

Abstract: The Educational mining is the subject to suggest right path to the students to get right carrier in their life. The aim of this paper is to discover of new patterns in higher education. The number of students did not know about the right progress in their higher studies .The lot of data mining approach could be used to solve this problem, likewise this paper tried to discover new patterns in higher education from hidden data using decision tree algorithm .the decision tree approach to classify by the records. The decision tree algorithm is used to predict number of chances in the higher education field in various disciplines .Predicting interest of students in a particular field is also important. it is also a key factor of the educational mining.

Keywords: Educational mining; decision tree; higher studies sector rank; category

1. INTRODUCTION

Education system is one the main system which is progressing our society .The better higher education studies is required for growth and development of country. Professional education is very important in higher education. In real world, predicting the performance of the candidates is a highly challenging task. Data mining, which is the science of searching deeply into databases for information and knowledge retrieval, has recently developed new applications and emerging discipline. These Educational Data mining techniques are applied to interpret the information about the candidates and explore the hidden knowledge from their resumes. One of the biggest challenges that higher education faces today is predicting the right paths to the candidates. Institutions would like to know, who will get placement first and who must be educate first for example, there are number candidates could enrol in particular course programs, but the institution should select the candidates who must need to go to post graduation if the candidate have interested to do his post graduation and his job simultaneously ,the institutions must support to and motivate to continue the higher education in the same institutions for better solutions.

One way to effectively suggest the advice to these candidates, through the analysis and presentation of data, or decision. Data mining enables institutions to report the capabilities to uncover and understand hidden patterns in vast databases. These patterns are then built into data mining models and used to predict individual status in higher education progress with high accuracy. As a result of, institutions are able to allocate resources to more effectively. Decision tree in Data mining, give n the information necessary to take action to efficiently allocate resources with an

accurate estimate of how many candidates will take a particular course.

To establish appropriate typologies for the number of candidates, in this research use Decision tree, the powerful clustering algorithm. They first applied the algorithms to the general groupings identified above, with mixed results. The boundaries among sets were unclear and dispersed, and even after repeated testing on holdout datasets, as well as the removal of suspected outliers

2. PREVIOUS WORK

Thiyaga rajan⁽¹⁾ explains about data mining in higher education in his research. The discovery of hidden patterns in educational data is a promising research in Educational DataMining. The students achievement rate were reduced continuously is the major problem in higher education. To increase the success rate of students the early forecast technique will help the management to counsel the poor students at right time. To discover the new patterns from various data the data mining approach is widely used. Likewise here the data mining is used in educational field to extract hidden patterns. Classification is used to classify the records based on the preparation set and also it uses the pattern to categorize the new records. This paper aims to show the various techniques of Educational data mining that guides the management to take better action on students at risk.

Jing luan⁽²⁾ is Using data mining in his research paper, the college discovered a way to make its mailing more effective and increase alumni pledges, whilereducing mailing costs. This is best described using a concept called "lift." If 20 percent of alumni respond to a pledge

request, the college should concentrate on those 20 percent. If data mining can quickly identify potential donors by a ratio of two to four (correctly predicting two out of four who will donate), then the university can achieve results by mailing only to the indicated 40 percent of the alumni population, thus saving considerable time and money.

According to **Ghadeer S. Abu-Oda and Alaa M. El-Halees, Faculty of Information Technology, Islamic University of Gaza – Palestine**^[3] to study the student dropout in computer science major in ALAQSA University for the purpose of improving the current teaching procedures and education strategies. Technologically, we do not propose new methods like FB-growth or decision-tree. However, we only use the mature classification and approaches in this study, cannot present more competitive algorithms or improve the existing algorithms. The study finds that mastering “digital design” and “algorithm analysis” courses has a great affect on predicting student persistence in the major and decrease student likelihood of dropout. **Dina A. Aziz AlHammadi, Mehmet Sabih Aksoy**^[4] have described in their paper is to discuss the need for data mining in all levels of any educational institute. An effective data mining cycle was presented to clarify the educational process.

3. PROBLEM DESCRIPTION

The previous researches have described data mining in higher studies. There is no prediction about over all disciplines. Number of studies focus on similar groups not for all. Here In my research is going to describe that how to utilise minimum number of attributes to predict the chances in higher studies in effective way for all set of students. It is going to recommend to use the hidden pattern in data mining to discover proper data from the data ware houses . it would match the right choice to the student to select their higher studies according to his or her interest . decision tree algorithm would match each and every scope of the discipline like a flowchart. The students can easily finding out their scope. The rural students could not understand many things in existing system. But decision tree structure give deliberate structure to each and every student who applied higher studies with minimum number of variable in al large database.

4. OBJECTIVES OF STUDY

1. To reduce the number of variables in a large data ware house.
2. To use right pattern to extract from the hidden data
- 3 .To predict right choice to the students
- 4..To utilise decision tree algorithm to classify and all categories of students.

Decision tree structure

Decision trees is an approach which have set of rules that lead to a class or value. Decision tree algorithm is the most popular Data Mining technique. It is a flow-chart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of a test, and leaf nodes represent classes or class distributions, the topmost node in a tree is the root node . Decision Tree is method of constructing a model from a dataset in the form of a decision tree or (equivalently) a set of decision rules by using very powerful. A table could be created in which the higher education chance for each possible input combination was stored. For e.g., SECT (U) r RANK (1-200) CAT (GEN), the percentage of candidates who had actually got placement in history database was computed. If this percentage is greater than 95%, this combination was classified as “exell chance. This decision tree has stored as an adjacency list. As an example, the partial view of the decision tree given in here. This process is repeated recursively for all cases. For all iterations, the best attribute field would be added to the adjacency list. One main point here is that the attribute field names, the results etc are all treated as nodes and the list can identify the node only by its corresponding TYPE. The ID field stores the unique id number of the node, while the parent stores the id of the parent of node the node. As an example the partial view of the adjacency list in table may be interpreted as if branch=CS, sect==Rural, rank high and cat== sc there is excellent placement chance. When this work is implemented as a web application, the candidate would enter his search criteria through the user interface screen and then a query is constructed, which are passed to get particular attributes which are used to search the decision tree.

5. PREDICTION OF HIGHER STUDIES CHANCES USING DECISION TREE ALGORITHM

Decision algorithm is the best prediction technique to choose the higher education. The candidate who has the interest in the higher education must go for it. So, this study give suggestion to that using decision tree. They can get priority of their status. It will help to get post graduation of relevant topic. It may give advantage to get job. But, there must be a plan in E-governance how many of them should be graduated. All the people should get the job. According to the need of vacancies, the seat must be allotted to getting higher education.

Table 1: predictors for higher education chance:

Criteria	Predictors	Priority	Chance
Sector	Rural	High	yes

Rank	Pass	Low	No
Category	SC	High	yes
Sector	Rural	High	yes
Rank	>90	High	yes
Category	SC	High	yes
Sector	Rural	High	yes
Category rank	>90	High	yes
Sector	Urban	Low	No
Rank	Pass	Low	No
Category	OBC	Low	No
Sector	Rural	High	yes
Category	SC	Low	No
Rank	>90	High	yes

		Yes	No	Total
Chance	Sect	1	4	5
	Rank	2	2	4
	Cat	3	2	5
				14

$$E(\text{Higher studies, Chance}) = P(\text{Sector}) * E(3, 2) + P(\text{Rank}) * E(4, 0) + P(\text{category}) * E(2, 3)$$

$$= ((5/14) * 0.971 + (4/14) * 0 + (5/14) * 0.971) = 0.693$$

Information Gain

The information gain based on the decrease in entropy after a data set is break on attribute.

Constructing is all about finding attribute that return most higher information.

Step I calculate entropy of the target

$$\text{Entropy}(\text{higher studies}) = \text{entropy}(\text{higher studies}) - E(\text{chance to higher studies, higher studies})$$

$$= 0.94 - 0.6993 = 0.247$$

ENTROPY FOR HIGHER STUDIES

$$\text{Entropy}(\text{Higher studies}) = \sum_{i=1}^c p_i \log_2 p_i$$

$$\text{Entropy}(\text{Higher studies}) = \text{entropy}(5, 9)$$

$$\text{The adjacency} = \text{entropy}(0.36, 0.64)$$

$$= (0.36 \log_2 0.36) - (0.64 \log_2 0.64) = 0.94$$

Entropy using the frequency table of two attributes

$$\sum (T, X) = \sum_{C \in X} P(c) E(c)$$

TABLE :2 HIGER STUDIES NUMBER OF ATTRIBUTES

Higher Studies

Step 3

Higher Studies	
Yes	No
9	5

Choose attribute with largest information the decision node divide dataset by its branches and repeat the process for every branch.

Higher Studies		
Attributes	Yes	No
Sector	3	2
Rank	4	0
Category	2	3

Gain = 0.247

Example applying decision tree algorithm

Sector	Rural	High	Yes
Sector	Rural	High	Yes
Sector	Urban	Low	No
Sector	Rural	High	Yes
Sector	Urban	Low	No

Example of prediction

Rank	Average	High	Yes
Rank	Average	High	Yes
Rank	Average	High	Yes
Rank	Average	High	Yes

Chances of prediction

Criteria	Predictors	Priority	Chance
Category	OBC	Low	No
Category	SC	High	Yes
Category	OC	Low	No
Category	OBC	Low	No
Category	SC	High	Yes

Step 4 a branch all must be yes, there are no branches.

Sector	Rural	High	Yes
Rank	>90	High	Yes
Category	SC	High	Yes

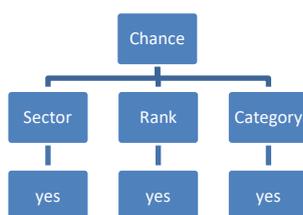
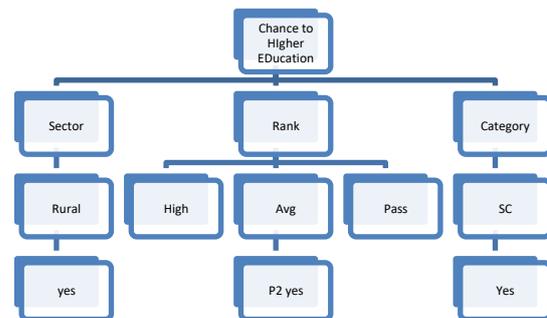


Figure prediction

The various predictors can give different results. The priority may be in the recurring procedures. If the priority must be given for the first position in the list,



the

Predictors make the important decision of the tree.

DECISION TREE DECISION ALGORITHM

If Sector = Rural AND Rank = Pass AND Category = High then Chance = yes

If Sector = Urban AND Rank = High AND Category = low, then Chance = yes

If Sector Urban AND Rank = Pass, AND Category = low, Then Chance = No

If Sector Rural, Rank = High AND Category = low, Then Chance = yes

If Sector = Urban, Rank = Avg AND Category = low, Chance = No

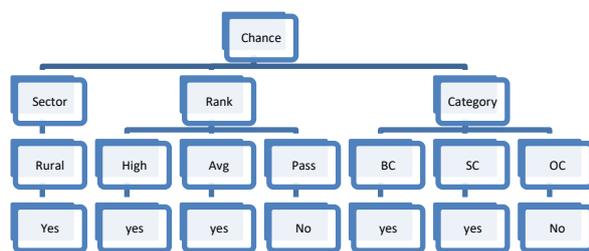


Figure prediction chances 1

6. CONCLUSION

The educational progress of candidates of any professional Institution has become the major issue of today. Predicting candidate interest in a particular discipline based on their Rank of test result components which are applying in the decision tree. Future work must include applying data mining techniques on an expanded data set which consider additional activities like diploma course and other vocational courses completed by candidates, which we believe may have a significant effect on the overall performance of the candidates.

REFERENCES

- [1] Dina A.Aziz AlHammadi Department of Information Sciences College of Computer and information Sciences King Saudi University P.O.Box 2454, Riyadh 11451 Saudi Arabia dina.alhammadi@gmail.com Mehmet Sabih Aksoy Department of Information Sciences College of Computer and Information Sciences King Saudi University P.O.Box 2454, Riyadh 11451 Saudi Arabia msaksoy@ksu.edu
- [2] Amirah Mohamed Shahiria, WahidahHusaina and Nuraini AbdulRashida – “A Review on Predicting Students Performance using DataMining Techniques” - The Third Information Systems InternationalConference, Procedia Computer Science 72(2015) 414 – 422
- [3] Heba Mohammed Nagy, Walid Mohamed Aly, Osama FathyHegazy –“An Educational Data Mining System for Advising Higher

EducationStudents” - International Journal of Computer, Electrical, Automation,Control and Information Engineering Vol:7, No:10, 2013

- [4] Monika Goyal and Rajan Vohra - “Applications of Data Mining inHigher Education” - IJCSI International Journal of Computer ScienceIssues, Vol. 9, Issue 2, No 1, March 2012
- [5] U. K. Pandey and S. Pal - “A Data mining view on class room teaching language” International Journal of Computer Science Issue,Vol. 8, Issue 2, pp. 277-282, ISSN:1694- 0814,2011
- [6] Dr. MohdMaqsood Ali - “Role of data mining in education sector” -International Journal of Computer Science and Mobile ComputingVol. 2, Issue. 4, April 2013
- [7] Dr. P. Nithya, B. Umamaheswari, A. Umadevi – “A Survey onEducational Data Mining in Field of Education” – International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 1, January 2016
- [8] KalpanaRangra, Dr. K. L. Bansal-“Comparative Study of Data Mining Tools” -International Journal of Advanced Research in ComputerScience .