

Predicting Student's Coming Target after Completion of Their Graduation by Mining Trained Database using Decision Tree

Sanjeev Gour, Apoorva Joshi

Abstract Data mining techniques have great capability to discover or find hidden knowledge from large data. As every educational institute produce large data about various academic and non-academic activities of their students in every year. This is why many emerging data mining techniques are widely used in educational domain to extract useful knowledge from huge database and this knowledge are then used for various educational decisions. Educational institute can predict or understand the academic and social behaviors of their students by mining their past data. This study, similarly emphasize on predicting student's coming academic or career target after completion of their graduation by analyzing past database which is generated by filling by them at the time of admission. This prediction can be used to understand the future behavior of the student so that educational institute could make better decision policy. One of data mining technique called decision tree is applied on trained database of students which is collected from department of computer science of Career College Bhopal using Rapid Miner tool. Many interesting and useful rules are extracted which help to make better decision policy in an educational institute.

Keywords: Decision Tree, Educational Data Mining, Prediction Analysis, Rapid Miner.

I. INTRODUCTION

Data mining techniques nowadays widely used by many educational institutes to analyze big database of their enrolled students. As many academic and extracurricular activities related to students are conducting throughout the session in a higher educational institute, they have stored a large unstructured data about their students during one academic year. This database increases day by day and become complex to analyze [1]. This is why today many data mining tools are used to analyze big educational database. This type of application of data mining is known as Educational data mining (EDM). EDM is one of emerging technologies used to extract useful knowledge or find hidden information from the large student's database. This knowledge and information are then used to make better decisions regarding many educational policies [2]. Classification and Prediction using decision tree are most popularly methods widely used in Education domain. Prediction data mining can be used to understand student behavior, predicting student's coming targets and outcomes. This study also used Prediction methodology using decision tree to predict or analyze student's behaviors after completion of their graduation by mining past database.

The study has completed by an experimental research which was implemented by one of data mining tool called Rapid Miner. The results of this experiment are interpreted and concluded in this study. More than 100 student's data of Career College Bhopal is used for this study.

II. PREVIOUS STUDIES IN THIS FIELD

As stated in introduction section that many educational institutes nowadays produce large database about their students so they are using many data mining methods to discover and extract useful knowledge from past database. This knowledge is then used to understand student's academic and social behaviors. Numbers of studies have been successfully implemented in educational domain. Author has put summary of some related studies in this filed here. In 2011, R.R. Cabra et al generate a prediction model using decision tree which predicts student's academic performance of an engineering college. The model is generated with help of popular mining tool called Weka [3]. Kolo David et al are also used decision tree approach in their study to predict student's academic performance using SPSS tool in 2015. They have concluded that factors about students such as Gender, financial status, previous grades and motivational level etc can affect the performance of the students in college [4].

Mashaël A. et al have identified the final grade of student by mining student's previous year grade for mandatory course in their research. For this study, authors have used a database of more than 200 students of King Saud University Riyadh in the year 2012. They have used classification and prediction algorithms in-built in the Weka tool [5].

In December 2018, Sanjeev Gour, in his recent study found some useful relationship among various students' personal attributes such as Gender, Category, Father-Occupation etc. with their interest in sport activity. Sport policy maker in an educational institute can use these extract relationships about the sport interest of students to make better decisions in sport framework. Author has used Weka tool again for this study [6]. Cédric Beaulac et al recently in august 2018, build classifiers which predict student's examination performance by mining previous year's grades using Random forest classification model [7]. Rahila Asif et al have identified courses which are indicators of good and poor performance of students in a particular course in their case study in 2014. This case study is implemented practically on more than 300 hundred students of NED university of Pakistan using Rapid Miner Tool [8].

Revised Manuscript Received on June 05, 2019

Dr. Sanjeev Gour, Associate Professor in Department of Computer Science in Career College, Bhopal, India.

Apoorva Joshi, Research Scholar in RNTU University Bhopal, India.

Predicting Student's Coming Target after Completion of Their Graduation by Mining Trained Database using Decision Tree

Many other such successful studies related to this filed motivate author to conduct this experimental study using decision tree which is implemented by mining tool Rapid Miner.

III. DATABASE FOR STUDY

The Dataset for this study was collected from the admission cell of Career College Bhopal. The dataset include some of main variables of student like Student's Course, Gender, Category, Scholarship, Father-Occupation, and student's choice in various coming target field like Higher Study, Placement, Entrepreneur, and Competitive Examination. The database was student admitted or enrolled in the session 2016-17 of department of computer Science of Career College. The data has been generated by college's local admission software at the time of student online registration process. For the study author has examined 9 major attributes (Table-I) of student. The data of more than 100 students were collected in Excel format. The Original format of dataset with selected variables has been shown in Table-I.

Table I- Details of Variables /Attributes Of Student's Database

SN	Name of Variable/Attribute	Role	Type
1	Course	Regular Variable	Polynomial
2	Gender	Regular Variable(Id)	Binominal
3	Category	Regular Variable	Polynomial
4	Scholarship	Special Target Variable (Lable)	Binominal
5	Father-Occupation	Regular Variable	Polynomial
6	Higher Studies	Special Target Variable(Lable)	Binominal
7	Placement	Special Target Variable(Lable)	Binominal
8	Entrepreneur	Regular Variable	Binominal
9	Competitive Examination	Special Target Variable(Lable & Weight)	Binominal
10	Research	Regular Variable	Binominal
11	BPL Card Holder	Regular Variable	Binominal

IV. METHODOLOGY

In this study, Classification also called supervised learning is used to classify or predict the target class label. The trained database is labeled. Some of main variables as their choice in student's database like higher study, Placement, Competitive Exam. Research and Entrepreneur are set to be target class label. Decision tree as a classifier classify or predict categorical class label and generate a tree model shown in Fig-III .In this study, Binary Variable Decision Tree is used as the target variables have binary categorical values such as YES or NO. The proposed methodology is shown in Fig.-I. Trained database is processed before passing to decision Tree classifier. Preprocessing of

database is done with the help of Transformation and Data cleaning operators of Rapid Miner tool. Preprocessed data are now ready to pass to the two classifier called Decision tree and Random Tree. Random Trees is nothing but set of individual decision trees where each tree is build from different class label and subsets of the training database.

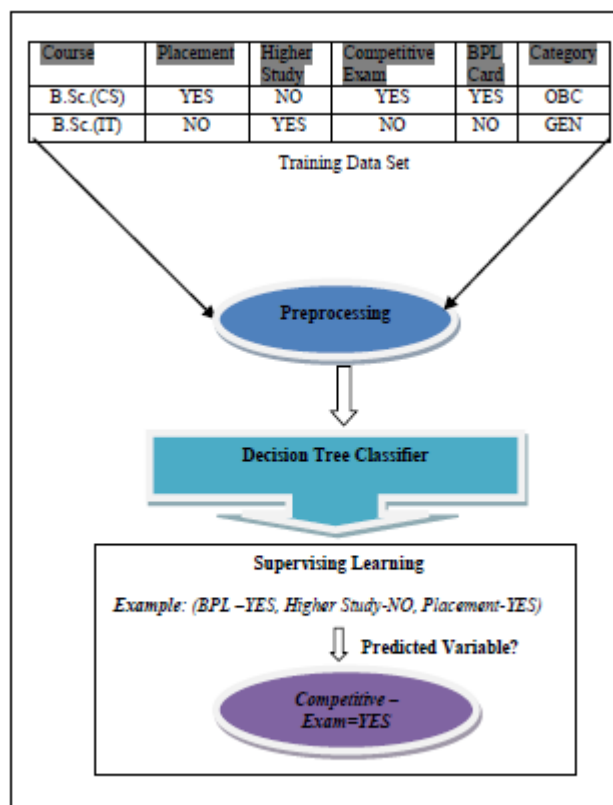


Fig. 1- Proposed Methodology

V. MINING TOOL

Rapid Miner Studio is a popular unified platform for data mining and predictive analytics in any data centric domain. It provides an easy visual design environment for making various analytics models with their built in machine learning algorithms. It has a huge collection of various machine learning algorithms, data preparation, exploration functions and visualization tools for analysis of various kind data. Rapid miner supports all kind of data mining procedures so this is why author used the latest version of Rapid Miner Studio for his analytical study.

VI. EXPERIMENTAL PROCESS

Step 1-Retrieve the Database: Browse the Student's Database on Mining toll from the local disk in Excel format.
Step 2-Transformation: Transform the format of database originally retrieved local disk into format more compatible with data mining process and tool used for this study.

Step 3- Set the Role and Type: Set the role and type of each variable considered for the study. Here author has selected different variable as a target variable (label) for each cycle of practical process. Set the type as polynomial, binominal, label, weight, id etc for each variable.

Step 4- Preprocessing: Set the default/replace values and parameters for missing, duplicate values and assign required values for error handling in database.

Step 5- Select Process: Select the Decision Tree and Random Tree processes from the prediction and classification model Operator for different target variables.

Step 6- Tree to rule: Select also Tree to Rule process for converting the results obtained by Decision tree into rules so that decisions for predicted or classified variable can be easily interpreted.

Step 7- Run the Process: After designing and making appropriate connections among applied processes mentioned in above first to sixth steps, execute or run the designed prediction model.

The stepwise experimental process with appropriate connection among applied process is shown in Fig -II.

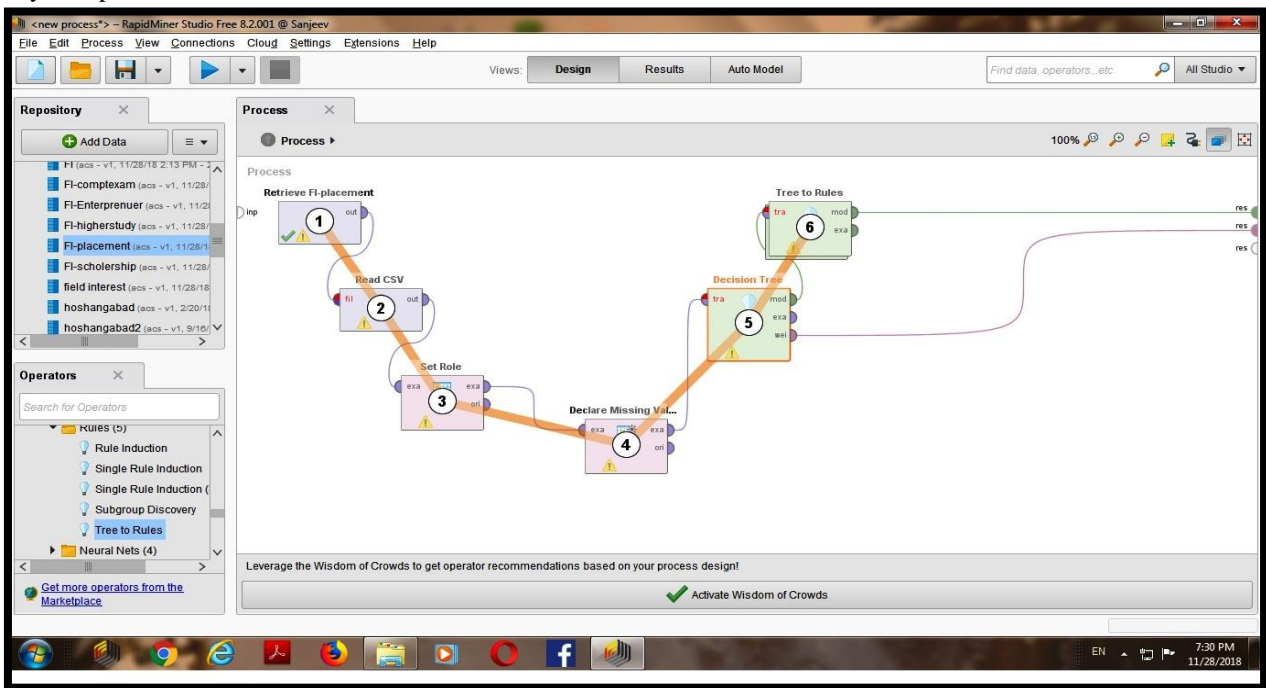


Fig. II- Screenshot of Design View for Experimental Process in Rapid Miner

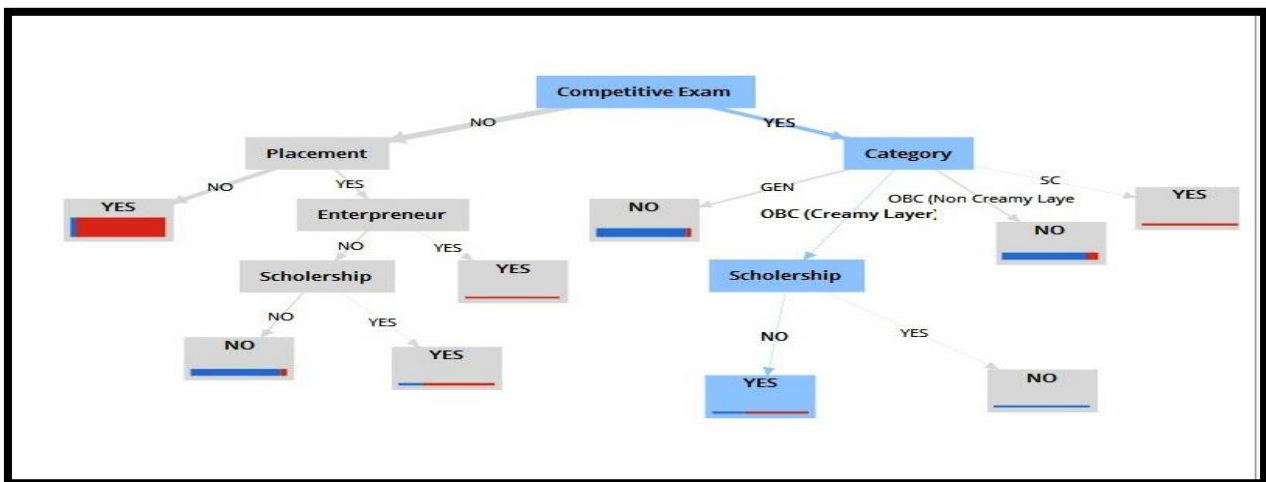


Fig. III- Decision Tree Generated For Target Variable “Higher Study” In Rapid Miner

Predicting Student's Coming Target after Completion of Their Graduation by Mining Trained Database using Decision Tree

VII. RESULT INTERPRETATION AND DISCUSSION

Different Decision trees are generated for different class labeled attribute. These trees are converted into decision rules which predict or classify the other data into target class. Following section describe some of useful and interesting classification rules with respect to different target variable.

Target Variable: Placement

Students belong to General class with possessing BPL card are not interested to go for higher study and are classified or predicted into placement cell while students belonging to OBC class with same options are not interested to go for placement. Students who don't have BPL cards are not classified into placement cell. This rule has more than 85% prediction accuracy.

Target Variable: Higher Study

Students whose father's occupation is private job are neither interested to become an entrepreneur nor go for competitive exam preparation nor not opted higher study while students whose fathers are farmers, opted higher study with same options. Students belonging to OBC class and preparing for competitive examination not interested to go for higher education while Students belonging to SC class, supposed to go for higher education with same options.

Target Variable: Competitive Examination

Male students who are having BPL cards, soundly interested in preparing for various competitive examination with continuing their study in higher education in comparison to female students. Students which does not come under BPL and studying in B.Sc. (Computer Science and Information Technology) are not interested in preparation for competitive examination.

Target Variable: Scholarship

More than 85% of students who are belonging to OBC (Non-Creamy Layer) class and their father working in private job or farming only, taking Graduation Scholarship from the Government.

Only male students whose father are working in private job and belong to OBC (Creamy Layer), applied for graduation scholarship.

Regular Variables: Research and Entrepreneur

Here it is one of noticeable point that not evens a single student wish to go in research area in his/her subject after completion of graduation and only 10 % students are interested to become an entrepreneur.

VIII. CONCLUSION

In this study, one of classification technique of Data Mining named Decision Tree is used on dataset of 120 students to predict and analyze student's coming academic target after completing their graduation. This study, found many useful predictions about student's interest by mining their past

data. Results as rules generated by Decision tree indicating whether a student likely to join higher study, competitive examination, research or will appear in placement conducted by college or to become an entrepreneur after completing their graduation. It is very obvious that if an educational institute has prior knowledge and information regarding student's interest about their future academic target, it can make better educational policy for their students.

IX. FUTURE ENHANCEMENT

Student's academic and non-academic behaviors can be analyzed by other attributes or variables of student's database like student's interest or choice in sports, cultural, social activities. These activities will surely affect the student's behavior. Data mining technique such as regression can also be used for numeric database for prediction analysis.

ACKNOWLEDGMENT

I have to express my appreciation to authors mentioned in following reference section for referring the part of their reports, surveys, observations, book and research manuscripts during the study of this research although any errors are my own and should not tarnish the reputations of these esteemed professionals.

REFERENCES

1. CH.M.H.Sai Baba, Akhila Govindu , Mani Krishna Sai Raavi , VenkataPraneethSomisetty, "Student Performance Analysis Using Classification Techniques", International Journal of Pure and Applied Mathematics, Volume 115 No. 6 2017, pp.1-7.
2. Ghadeer S. Abu-Oda and Alaa M. El-Halees, "Data Mining In Higher Education: University Student Dropout Case Study", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, January 2015, pp.15-27.
3. R.R. Cabra, R.S. Bichkar," Performance Prediction of Engineering Students using Decision Trees", International Journal of Computer Applications (0975 – 8887) Volume 36– No.11, December 2011 pp.8-12.
4. Kolo David Kolo, Solomon A. Adepojob, John Kolo Alhassan, " A Decision Tree Approach for Predicting Students Academic Performance" , I.J. Education and Management Engineering, 2015, 5, pp.12-19.
5. Mashael A. Al-Barrak and Muna Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study ", International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016, pp.528-533.
6. Sanjeev Gour ," Finding Relationship among Various Students' Attributes with Sport Interest using Association Rule Mining Method" , Asian Journal of Computer Science and Technology ISSN: 2249-0701 Vol.7 No.3, 2018, pp. 72-75.
7. Cédric Beaulac Jeffrey S. Rosenthal, "Predicting University Students' Academic Success and Major using Random Forests " , 23 august,2018 pp.1-23.
8. Raheela Asif, Agathe Merceron, Mahmood K. Pathan," Predicting Student Academic Performance at Degree Level: A Case Study", I.J. Intelligent Systems and Applications, 2015, 01, pp.49-61.

AUTHORS- PROFILE



Dr. Sanjeev Gour pursued Bachelor of Computer Science (Honors) and Master of Science (Electronics & Communication) from Devi Ahilya University Indore and Master of Science (Computer Science) from Barkatullah University Bhopal. He is completed his Ph.D. in Computer Science and currently working as a Associate Professor in Department of Computer Science in Career College, Bhopal. He was a member of Board of Studies & Examination Committee of Computer Science in Barkatullah University Bhopal. He is a member of Managing Committee in Computer Society of India (Bhopal Chapter 2017-18). He has published more than 15 research papers in - reputed international journals including Thomson Reuters & Scopus (SCI & Web of Science) and presented more than 20 research papers in national and international conferences. His main research work focuses on Data Mining. He has 16 years of teaching experience and 7 years of research experience.



Apoorva Joshi is working as a Research Scholar in RNTU University Bhopal. She is having Six years' experience as an Assistant Professor & in the research field. Her Ph.D. is in Date Mining Privacy area. Her Educational qualification is MCA. Her areas of specialization are Data Mining, Data Privacy & Security, Big Data, and Privacy- Preserving & Information Retrieval.