

# A Data Mining Model to Improve Placement

Ravi Tiwari  
Department Of Computer Science And  
Engineering,  
M.M.M.U.T, Gorakhpur

Awadhesh Kumar Sharma, PhD  
Department Of Computer Science and  
Engineering,  
M.M.M.U.T, Gorakhpur,

## ABSTRACT

Education data mining is one of the growing fields of the present time. as it grows many issues to improve system comes in the notice one of them is improvement of the placement. Placement is a very important issue for any educational organization. Every organization wants to improve its placement. Success of any educational institute is measured by the placed student of the organization. This paper actually deals with the application of neural network to the educational data to improve placement. It takes education data i.e student data as a input and predict the status of placement for the student. it also suggest the best field according to nature of the student data for particular student which have a chance to place. So through this educational organization can predict the placement of each student and can work according to it to improve the placement of the organization.

## Keywords

Educational data mining, predication, random tree .

## 1. INTRODUCTION

Students placement is one of the important thing for any educational institutes because it is one of the parameter which actually defines the success or failure of any organization. a good placement leads to good reputation As well as the good future of the organization. Beside this placement predication also helps the organization to find the weak student in the organization and helps the organization to focus on the weak students. This thing attracts so many researchers to work in this field.

Data mining techniques are widely used for the prediction purpose in education sector. In this paper we use the data mining process to explore the various parameters which affect the student placement. Our paper uses random tree technique for the prediction of student placement. In this paper we consider the students batch 2013 of B. tech and collects data related to them regarding to their academic performance, participation in the various cultural activities and their current placement status and if placed then what are they doing at present. We use this information to build a model using random tree algorithm technique which not only predict the status of the placement but also suggest which type of profession will be best for the student if its status is placed.

## 2. BACKGROUND AND RELATED WORK

Before building this model we have gone through the work done by various researchers. Romero and Ventura[1] done a lot of important work related to education data mining. It uses data mining techniques for the purpose of predication in education field. Nguyen et al. [2] compares the accuracy of decision tree and Bayesian network algorithm for the

prediction of performance of the undergraduate and post graduate students. Results from this work are useful for prediction of the weak student who may fail in the exam. Affendy and Musthain [3] uses performance in various subjects to predict the CGPA of bachelor students. Al-Radaieh et al.[4] uses classification technique to improve the quality of education. Cesar et al [5] uses data mining technique to develop a model which helps student to take academic decision. Nghe et al also provides the a lot of contribution in this field. Ramaswami and Bhaskaran also develop a predictive model to evaluate achievement of student at higher secondary level. N.S.Shah applies various decision tree technique to categorize bba student based on performance. Tripti mishra et al [6] also uses classification technique to predict student's performance. In our paper we use random tree algorithm to develop a model to predict placement status for a student and actually contributes to the overall placement of the Institution.

## 3. DATA COLLECTION

Data collection is a very important step of building prediction model because it is the base of all your predictions, a minor error in the data cause a blunder in the prediction. so data collected must be accurate. In this model we collect the data related to engineering undergraduate students of the batch 2013. Data collection involves the data related to the various attributes which are considered in the model for predictions. The attributes consider in this model are gender , category , student marks in the first second and third year , the confidence level of student , students native place from which he or she belongs, student family background i.e. rich or poor, leadership ability, student's participation in the extracurricular activities , interaction of student within the social group.

Attributes	values
Gender	Male (0), female (1)
Category	GEN(1) , others (0)
First year	Marks (grade)
Second year	Marks (grade)
Third year	Marks (grade)
Communication	0-10
Belongs	City (0) , village(1)
Family background	rich(1) , average and poor(0)
Leadership	0-10
Participation in extracurricular activities	0-10
Interaction	0-10

#### 4. BUILDING PREDICTION MODEL

After the collection of desired data we use Weka data mining tool for the development of prediction model using random tree algorithm. We simulate the random tree algorithm on the data of b.tech students of batch 2013.

For applying the various algorithm on the given data set we need to convert excel file to arff file . For converting Excel file to Arff we use the note pad. By doing this we get the required file for processing in weka.

```
latest.arff - Notepad
File Edit Format View Help
@relation student_data
@attribute gender { m , f }
@attribute category {gen ,cate }
@attribute first {A, B,C,D,E,F,G}
@attribute second {A, B,C,D,E,F,G}
@attribute third {A, B,C,D,E,F,G}
@attribute communication { A,B,C,D}
@attribute belong {C,V}
@attribute family_back {R,P,M}
@attribute leadership {Y,N}
@attribute participation { A,B,C}
@attribute interaction { A,B,C}
@attribute job {gov,IT , NA , research, management}

@Data
m,gen,B,B,A,B,C,M,N,B,A,gov
m,gen,D,F,D,B,C,P,N,C,A,NA
m,gen,C,E,C,B,C,M,N,A,A,NA
m,gen,B,B,B,A,C,M,Y,A,A,IT
m,gen,C,E,D,B,C,M,Y,A,A,IT
m,cate,E,F,D,C,C,R,N,C,A,management
m,cate,F,F,D,C,C,M,N,C,A,NA
f,gen,C,D,C,A,C,M,Y,A,A,IT
m,cate,F,G,D,C,C,M,N,C,B,NA
m,cate,D,D,C,B,C,M,N,A,A,research
m,cate,F,G,E,C,C,R,N,B,B,management
m,cate,E,E,C,B,C,M,N,A,B,IT
f,gen,C,B,B,C,C,M,N,B,B,research
m,gen,E,F,D,C,C,R,Y,B,B,management
m,gen,A,C,A,B,C,M,N,A,A,IT
f,cate,D,E,B,B,C,M,N,A,B,IT
f,gen,B,C,B,B,C,M,N,A,A,IT
m,gen,A,B,A,B,C,M,N,A,A,IT
f,gen,C,C,B,A,C,M,Y,A,A,IT
m,gen,B,C,A,B,C,M,N,B,A,IT
```

Fig1. Model for the prediction process

Now we apply various algorithm on the student information data set using weka tool.

We apply ID3 , BayesNet , J48 , RBF network ,Random tree algorithms on the student information data set using weka tool.

#### 5. RESULT AND ANALYSIS

By apply ing the various algorithms on the given dataset we find following results .

As figures shows after applying the various algorithm on the student information data set we find Id3 Fig1 is 71% accurate , BayesN et is 70% accurate , J48 is 71% accurate , RBF network algorithm is 65 % accurate and Random Tree algorithm is most accurate for the prediction process . i.e 73% .

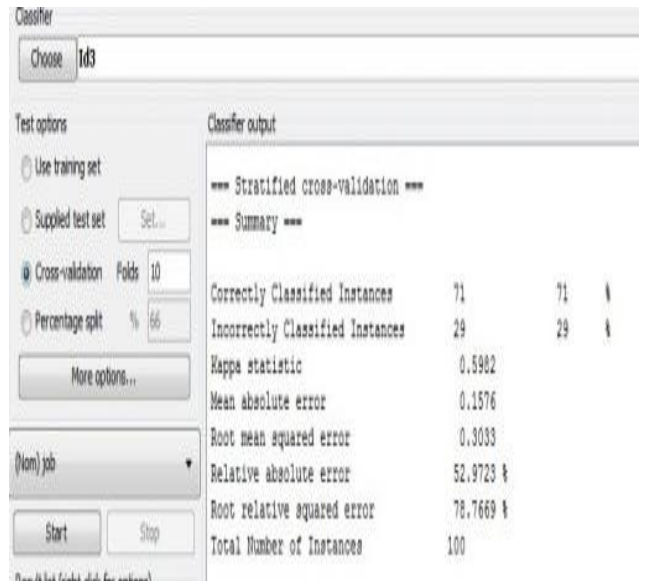


Fig1. Applying Id3 algorithm

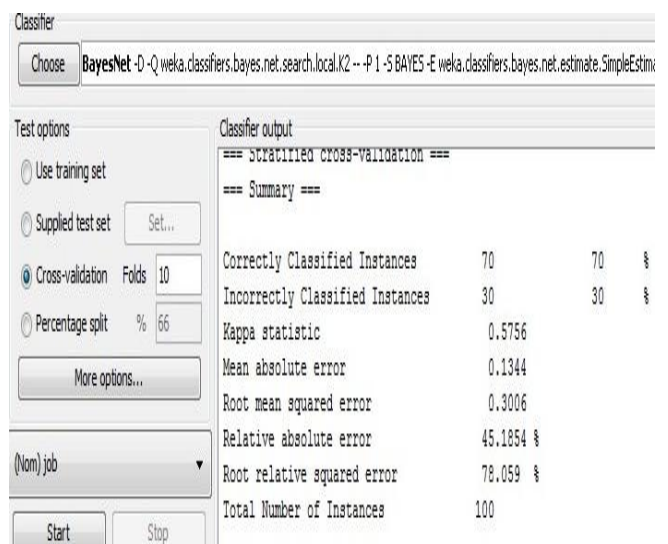


Fig2. Applying BayesNet algorithm

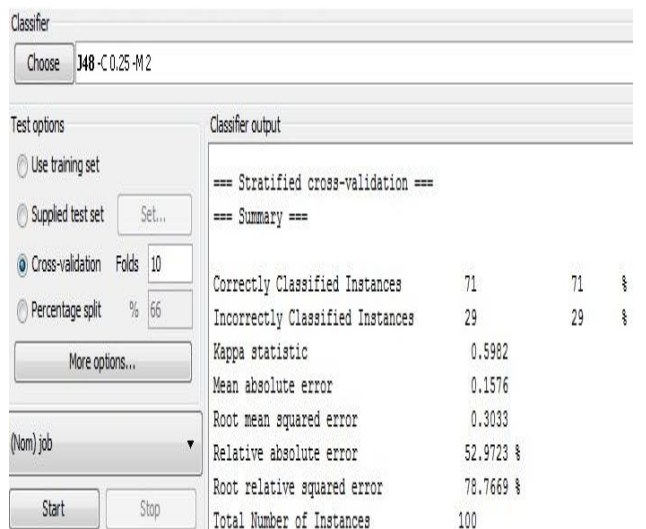


Fig3. Applying J48 algorithm

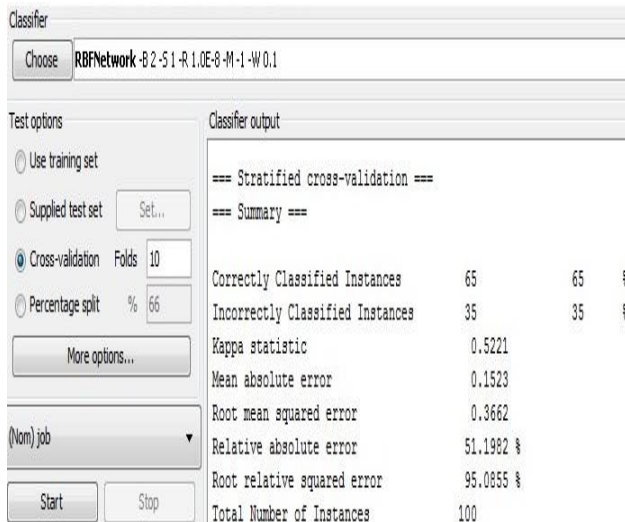


Fig4. Applying RBF Network algorithm

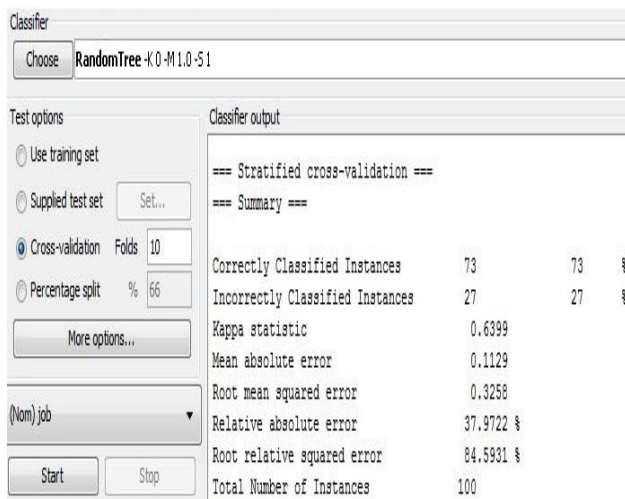


Fig5. Applying Random Tree algorithm

So Random tree algorithm is best suited for prediction of placement status for a student. Therefore by using Random tree algorithm we generate a decision tree which can be used for the building the model for predicting placement of students accurately.

## 6. CONCLUSION

Today placement is one of the critical thing for the educational institutes. So prediction of student's placement can help them to provide assistance to the needy student to improve the overall placement of the college. It also helps in the development of the system which actually suggest to students in selecting career.

## 7. REFERENCES

- [1] Romero, C., Ventura, S. and Garcia, E., "Data mining in Course management systems: Moodle case study and Tutorial". Computers & Education, Vol. 51, No. 1.pp.368- 384. 2008.
- [2] Nguyen N. , Paul J. , and Peter H. , a comparative analysis of techniques of predicting student performance. In the proceeding of 37<sup>th</sup> ASEE/IEEE frontiers in education.
- [3] I.H. , M.Paris , L.S.Affecndy , improving prediction performance using voting technique in data mining , in world academy of science, vol 38 , 2010
- [4] Al- Radaideh , Al-Shwakfa , mining student data using decision tree ,in international Arab conference on information technology , 2006
- [5] Cesar V. , Javier B. , liela S. ,recommendation in higher education using data mining , educational data minng conference 2009
- [6] Tripti mishra , Dr. dharminder kumar , dr.sangeeta gupta , mining student data for performance prediction , fourth international conference on advance computing and communication technologies, 2014