International Journal of Engineering & Technology, 7 (1.2) (2018) 43-46

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Educational data mining for student placement prediction using machine learning algorithms

K. Sreenivasa Rao [1]*, N. Swapna [2], P. Praveen Kumar [3]

[1] *Department of Computer Science and Engineering, Vignana Bharathi Institute of Technology, Hyderabad, India*
ksrao517@gmail.com

[2] *Department of Computer Science and Engineering, Vignana Bharathi Institute of Technology, Hyderabad, India*
swapnakiran29@gmail.com

[3] *Research Scholar, Veltech Dr.RR & SR University.Chennai,India*
praveen.padigela@gmail.com

**Corresponding author E-mail: ksrao517@gmail.com*

Abstract

Data Mining is the process of extracting useful information from large sets of data. Data mining enablesthe users to have insights into the data and make useful decisions out of the knowledge mined from databases. The purpose of higher education organizations is to offer superior opportunities to its students. As with data mining, now-a-days Education Data Mining (EDM) also is considered as a powerful tool in the field of education. It portrays an effective method for mining the student's performance based on various parameters to predict and analyze whether a student (he/she) will be recruited or not in the campus placement. Predictions are made using the machine learning algorithms J48, Naïve Bayes, Random Forest, and Random Tree in weka tool and Multiple Linear Regression, binomial logistic regression, Recursive Partitioning and Regression Tree (rpart), conditional inference tree (ctree) and Neural Network (nnet) algorithms in R studio. The results obtained from each approaches are then compared with respect to their performance and accuracy levels by graphical analysis. Based on the result, higher education organizations can offer superior training to its students.

Keywords: *Data Mining, Educational Data Mining, machine learning algorithms.*

1. Introduction

The purpose of higher education organizations is to offer superior opportunities to its students. Placements are considered to be very important for each and every student in the college. Colleges are opted by parents and students based on placement record of the organization, Organizations are ranked based on placement record. Hence it is beneficial for every organization to have an approach of predicting the placement chances of each student based on some attributes and parameters.

Educational data mining involves new methods and approaches for discovering the knowledge by analyzing the student databases to support the decision making process in educational institution in offering the best training to their students.

The placement of a student not only depends on his/her academic capabilities but also involves the attributes such as performance in placement assessment examinations conducted by assessment agencies (ex.co-cube), communication skills etc. and thus decisions are made towards the best prediction in the campus placement and also which parameter of the student is contributing more towards placement of the student. In this work we collected final year student data comprising of 5 attributes SSC %, Inter %, B.Tech aggregate % co-cube score an attribute called "placed" that tells us whether the student got placed or not ?. Machine Learning algorithms are applied in weka tool and R studio on final year student dataset. Actual and predicted placement status is compared for accuracy. The efficiency/accuracy of each model is visualized and tested and based on the performance analysis, each model results are discussed.

2. Literature survey

Molina et al. presented a case of study with educational datasets using Meta-learning approach for automatic parameter tuning [1]. They used 14 educational data sets and J48 algorithm with only 2 of its parameters and concluded that meta learning approach can be used for parameter tuning of decision tree algorithms.

T. Jeevalatha, et.al used the decision tree algorithm to predict the selection of student for the placements [2]. They used Decision Tree (DT) algorithm such as C4.5, ID3, and CHAID which were developed by using Data Mining Rapid Miner software/tool. Neelam Naik and Seema Purohit built the model to classify the performance of the placement of students [3]. The error produced to classify validation data, result prediction classification tree was 38.46% and while for validating placement prediction classification tree was found 45.38% respectively.

Ajay Kumar Pal and Saurabh Pal collected the data for the study and analysis of the student's educational performance basically for training and placement. The authors used different classification

algorithm and used WEKA data mining tool [4]. They concluded that naive Bayes classification model is the better algorithm based on the placement data with found accuracy of 86.15% and overall time taken to build the model is at 0 sec. As compared with others Naïve Bayes classifier had lowest average error i.e. 0.28.Ajay Shiv Sharma, and et.al, used the logistic regression model and developed the placement prediction system (PPS) [5]. The accuracy of training and testing of the algorithm was 98.93% and 88.333%.

BahenSen, EmineUcar and DursunDelen collected the large and feature rich dataset and build the model to predict the placement test results [6]. They used support vector machine, C5 Decision Tree algorithm, and artificial neural network. They resolved that C5 Decision Tree algorithm is the better prediction model with efficiency of 95%, the accuracy of support vector and artificial neural network is 91% and 89%. Mangasuli Sheetal B and Prof. Savita Bakare made predictions using the Data Mining Algorithm "Fuzzy logic" and "K nearest neighbor (KNN)" [7]. The accuracy obtained after analysis for KNN is 97.33% and for the Fuzzy logic is 92.67%.

Ramanathan.L et al. predicted the placement of student by using similarity measure with mathematical method which is called sum of difference (SOD) [8]. They made it obvious that placement is not so easy to predict because it depends on many attributes, even the paper is considered with four attributes.

Hari Ganesh et al. [9] discussed various applications of Data Mining. They summarized various data mining techniques, algorithms their contribution to various areas of Educational Data Mining. They concluded that apart from contribution of EDM in higher education, EDM can be extended to analyze knowledge process of primary class students to know their learning problems.

John Jacob et al. [10] predicted student performance using data mining techniques like Regression and decision trees to know academic failure of students. They also used clustering to group the students as per their academic performance based on their strengths and weaknesses. They identified that apart from challenges and cost involved in EDM, EDM implementation requires privacy and ethics of all the stakeholders involved in EDM process.

3. Methodology

Proposed placement prediction system is equipped with various data mining tasks and is depicted in the following Architecture diagram. Educational data consisting of students' details and their marks etc. are collected. In our dataset we collected Serial no, HT.No, SSC %, B.Tech % and Co-cube score. Then preprocessing is performed to eliminate irrelevant attributes like Serial No. and HT.No since (because) they do (will) not play any role in analysis. Entire dataset is replicated into two sets namely training data and test data, there is no difference between two sets except predicted placement status is placed in one column before the original placement status column. Machine algorithms like J48, Naïve Bayes, Random Forest, and Random Tree are performed in weka tool to build the model and the learned model is used to predict the placement on test data. After predicting the placement on test data, we compare the actual placement and predicted placement for accuracy analysis. Similarly, algorithms namely Multiple Linear Regression, binomial logistic regression, Recursive Partitioning and Regression Tree (rpart), conditional inference tree (ctree) and Neural Network (nnet) algorithms are also applied in R studio and placement is predicted.

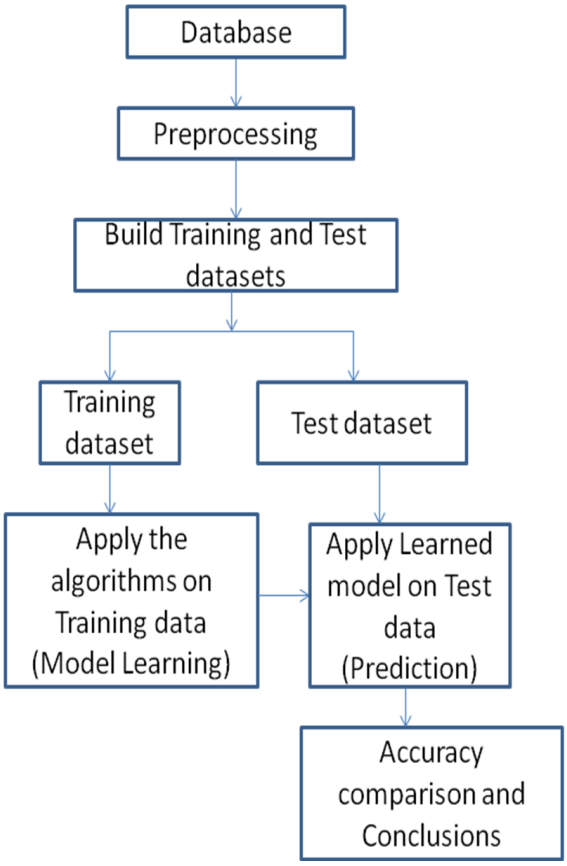

Fig. 3.1: Placement Prediction System Architecture.

4. Experimental results discussion

4.1. Performance of algorithms in weka tool

We performed various machine algorithms in our student dataset in weka tool and tabulated the analysis parameters in the table given below.

Above table shows random forest and random tree algorithms are giving 100 % accuracy on student placement dataset and J48 algorithm has 88.89 % accuracy and Naibe Bayes got only 61.10 %. It depends on nature of dataset, since our data has only numericals, (so) random forest and random tree algorithms performed well. On other datasets J48 and Bayes may perform well.

Table 4.1: Performance of Machine Learning Algorithms on Student Dataset

	J48	Naïve Bayes	Random Forest	Random Tree
True Positives	14	11	23	23
False Positives	1	23	0	0
True Negatives	66	44	67	67
False Negatives	9	12	0	0
Accuracy	88.89%	61.10%	100%	100%

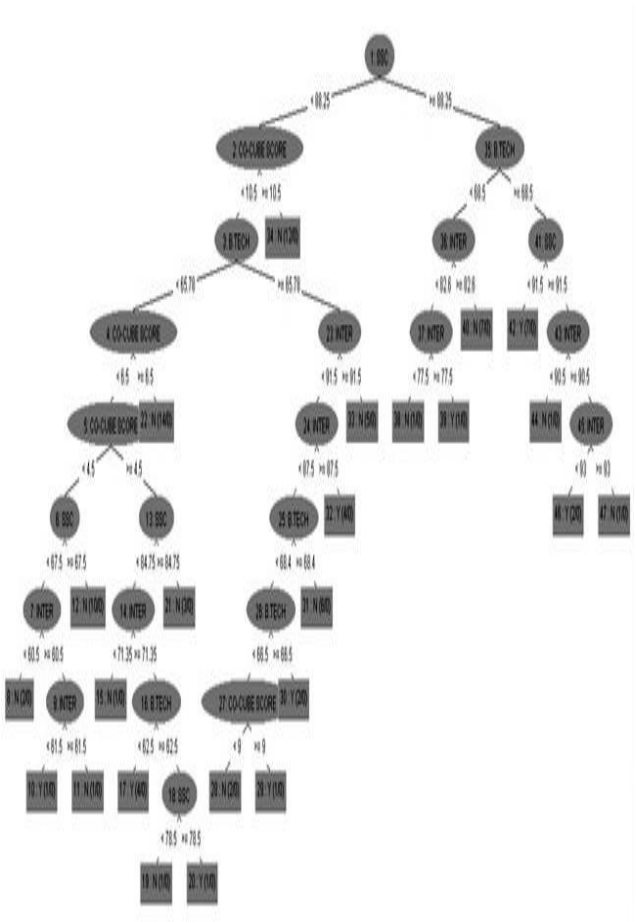

Fig. 3.2: Classification Tree Generated by Random Tree Algorithm.

4.2. Performance of algorithms in R Studio

We performed multiple regression on out dataset and analyzed which attribute of the student is more contributing towards placement of the student by eliminating and adding each of the attribute in multiple regression. Tabulation of regression output is also represented in Table 4.2.

From above table it is evident that B.Tech percentage has contributed more towards placement of the students whose data is collected in this paper. It is the insight made for this dataset only; it may be different for other datasets.

We also performed various machine learning algorithms in R studio like Multiple Regression, binomial logistic regression, recursive partitioning & regression tree, conditional inference tree and neural networks and tabulated the accuracy in table…, here it is obvious that recursive partitioning & regression tree offers high accuracy compared to other methods. Also performances of algorithms depend on nature of the dataset.

Table 4.2: Multiple Regression Summery on Dataset

Model	Variables	Significance	P-Value	Model P-Value	Adjusted R-Square	Remark
1	SSC B.TECH	.	0.08	0.06229	0.04	
2	B.TECH	*	0.0426	0.04263	0.035	MORE SIGNIFICANT
3	SSC B.TECH CO.CUBE	.	0.2114 0.0674 0.5516	0.1177	0.033	
4	B.TECH CO.CUBE	*	0.0414 0.6273	0.1151	0.0266	
5	CO.CUBE		0.706	0.706	-0.009	NOT SIGNIFICANT

Indicates nominal significane
* indicates significance

Table 4.3: Performance Analysis of Various Algorithms in R Studio on Our Dataset

	Multiple Regression	Binomial Logistic Regression	Recursive Partitioning & Regression Tree	Conditional Inference Tree	Neural Networks
True Positives		3	10	0	
False Positives		3	4	0	
True Negatives	67	64	63	67	67
False Negatives	23	20	13	23	23
Accuracy	74.44%	74.44%	90%	74.44%	74.44%

5. Conclusion

Here educational data mining is performed on final year student information of an organization to predict campus placements of students. Machine learning algorithms are performed in weka environment and R studio. Results of application of algorithms are tabulated and analyzed that shows random tree algorithm gives 100 % accuracy in prediction on our dataset and also in R environment Recursive Partitioning & Regression Tree performs better and gives 90 % accuracy. We also accept that performance depends on nature of dataset. We conclude that B.Tech percentage attribute is contributing more towards placements of students. Again it may vary from dataset to dataset.

References

[1] Molina, M. M., Luna, J. M., Romero, C., & Ventura, S., 2012, "Meta-learning approach for automatic parameter tuning: a case of study with educational datasets", in Proceedings of the 5th international conference on educational data mining, pp.180-183.

[2] T. Jeevalatha,N. Ananthi,D. Saravana Kumar, "Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms", International Journal of Computer Applications (0975 – 8887),Volume 108 – No 15, December 2014.

[3] NeelamNaik and SeemaPurohit, "Prediction of Final Result and Placement of Students using Classification Algorithm", International Journal of Computer Applications (0975 – 8887), Volume 56– No.12, October 2012.

[4] Ajay Kumar Pal and Saurabh Pal, "Classification Model of Prediction for Placement of Students", I. J. Modern Education and Computer Science, 2013, 11, 49-56.

[5] Ajay Shiv Sharma, Swaraj Prince, ShubhamKapoor and Keshav Kumar, "PPS - Placement Prediction System using Logistic Regression", IEEE International Conference on MOOC, Innovation and Technology in Education (MITE), 2014.

[6] BahaSen , EmineUcar and DursunDelen , "Predicting and analyzing secondary education placement-test scores: A data mining approach", International journal of Expert system with applications, Volume 3 , 2012,Issue 10, pgno: 9468-9476.

[7] Mangasuli Sheetal B1, Prof. Savita Bakare, "Prediction of Campus Placement Using Data Mining Algorithm-Fuzzy logic and K nearest neighbor ", International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 6, June.

[8] Ramanathan.L et al. "Mining Educational Data for Students' Placement Prediction using Sum of Difference Method", International Journal of Computer Applications (0975 – 8887) Volume 99– No.18, August 2014

[9] S.Hari Ganesh A.Joy Christy "Applications of Educational Data Mining: A Survey ", IEEE Sponsored 2nd International Conference ICIIECS-15.

[10] John Jacob, Kavya Jha,Paarth Kotak, Shubha Puthran "Educational Data Mining Techniques And Their Applications", IEEE International Conference On Green Computing and Internet Of Things (ICGCIoT),2015. https://doi.org/10.1109/ICGCIoT.2015.7380675.