Predictive Analytics for Placement of Student- A Comparative Study

Sonali Rawat

Assistant Professor, Department of Computer Science Chandigarh University Mohali, Punjabi

---

Abstract - *Predictive Analytics is science of withdrawing information from the existing data for the purpose of determining existing patterns and future trends. Scope of this review paper is to study different data mining techniques that can be used to analyze the academic performance of students to predict their chances of getting placed through campus placement. Different attributes such as academic results, technical skills, training and projects done, are considered to be desirable for prediction purpose. This paper presents an outlook to differentiate data mining techniques for predictive analytics that can be used in the process of predicting campus placement.*

***Key Words*: Predictive analytics, Data Mining, Classification, Decision Tree**

1. INTRODUCTION

Predictive Analytics presently has many applications including banking, health, social media, retail etc. Due to its positive result new areas are also emerging to use it, one such area is "EDUCATION". Predictive Analytics can be used in Educational Institute to predict placements of students as employability has become one of the crucial businesses in the present world. Monumental amount of students takes admission in professional college with the hope of acquiring their dream job. So, it would be a great deal if institute as well as student can get idea of placement beforehand .The result of prediction is informed to the students so that they can upgrade and refine their skill in order to increase the graph of recruitment.

2. Literature Review

Classification Approach: - Classification is used to classify each data item into one of the predicted target class or group and to accurately predict categorical labels. Classification uses classification models to predict the class label. Using a set of predefined classes, class label of each object is determined. Training set is provided as an input to algorithm to build model, which can be used for classification of new object. For example, a bank starts credit policy for his customers; manager by the behavior of customer can classify them under three categories: "safe", "risky", "very risky". So classification will help us to draw a model that could be used to accept or reject future request for the credits.

Sudheep Elayidom et al. [1] aim to use data mining technique for the benefit of students in future. They uses different techniques like decision tree, naive bayes and artificial neural networks and declare four class labels excellent, good, average and poor for each branch. A student needs to enter his entrance rank, gender (M/F), sector (rural/urban) and reservation category, and then using data mining techniques, he or she may know which branch is suitable for him or her. Then with the help of above information a student enters his branch, location etc. and on the root of which the placement chances for different streams of study is calculated. Hence student may opt for the branch providing chances of excellent placement. At the end of the paper the three techniques are compared and it shows that decision tree is slightly good in terms of accuracy however the difference is unimportant. And therefore there is no universally accepted best model.

Tripti Mishra et al. [2] provides classification technique like Bayesian method, multilayer perceptrons and sequential minimal optimization (SMO), Ensemble methods, decision tress using WEKA and emotional skill like assertion, empathy, decision making, leadership and stress management to predict placement of MCA students. ROC curve and F measure are used to compare these algorithms. Emotional skill parameters are assessed through Emotional skill assessment process (ESAP) tool. All the models are compared and **J48** is suggested as the best technique among all with the best accuracy and least time to build.

T.Jeevalatha et al. [3] used decision tree algorithm like ID3, CHAID, and C4.5 using **Rapid Miner tool** to predict student performance in placement activity using two year student records. All three algorithms are compared and the result shows that ID3 has highest accuracy rate of 95.33 percent.

Neelam Naik et al. [4] provides prediction to gain the knowledge about students of Master of Application before admitting them to the course. Sample of around 325 students is taken among which 195 records are used to create model for prediction of result and placement and remaining are used for validation purpose. Data mining algorithm are applies using XL miner tool and a classification tree is developed, on the basis of which decision rules are made. These rules are implemented using ASP.net

software. Error experienced during the experiment in the paper is 38.46% for MCA result prediction and 45.38% for placement prediction. A keen observation is made in the paper that classification tree do not take gender as a trait for taking decision.

Pratiyush Guleria et al. [5] uses application of Bayesian theorem to predict the result of placements. The traits used are attendance, GPA, reasoning, quantitative, communication, and technical skill. Two tools are used for the classification of the record – WEKA & RAPID MINER. At last the result is represented with the help of a graph which shows that student with good technical skill have comparatively more chances of getting placed.

Prof. Savita Bakare et al. [6] defines education data mining as the use of data mining technique in the process of education. Database of students is analyzed and studied using mining technique to predict whether student enlists during the campus placement. The attributes taken for prediction include academics as well as co-curricular activities along with communication skills etc. "Fuzzy logic" and "K nearest Neighbor (KNN) are used for building model. Data of 900 students is taken among which 600 are used as training set for model building and remaining as test data for validating the model. At the end result of both the method is calculated and observed that fuzzy has accuracy of 92.67% with running time of 450msec and KNN has accuracy of 97.33% with running time of 13458msec. So KNN is concluded as better method.

Revathy S et al. [7] focuses on the importance of Education data mining (EDM) technique which can be used to traverse concealed information of students from their resumes. The paper uses data mining techniques to get an idea of students composing for the coming placement activity. A reliable framework is designed to locate students to be placed in whole database. Classification technique is used to categorize student according to their academic documentation. The specifications used for categorization are academic detail, technical skill, programming skill, quantitative and reasoning skill. To forecast about the company student is likely to be placed. C5.0 algorithm is used for classification, which result in decision tree formation using Quinlan. The prediction is done using R, where data is divided into two parts one is training data other is test data. The output predicted is displayed using a pie chart and accuracy of 75% is observed.

Namita puri et al. [8] intent to design a model for the prediction of students placement by identifying appropriate traits based on academics skills and curricular of final year student. The model is designed using a classification technique based on decision tree. The paper suggests use of ID3 algorithm for the implementation. The result stipulates accuracy of algorithm with 95%.

Mosima Anna Masethe et al. [9] notifies that large databases of educational institutes can be studied with the help of data mining techniques to discover hidden pattern in it in order to use them for the purpose of decision making. Various classification algorithms like j48, simple cart, bayes net, naïve bayes and reptree algorithm are applied to database so as to predict placement in work integrated learning. Confusion matrix for all algorithms is formed and was observed that naïve bayes and bayes net gave best result.

Ravi Kumar Rathore et al. [10] targeted to use fuzzy inference system from Matlab tool box in order to differentiate the information of student based on their performance. Data of 31 students from M-tech CS including X th marks, XII th marks, B-tech cgpa, M-tech cgpa, and data of backlog is collected, then membership function is assign for each input set which shows that student with membership function greater than 2.1 is eligible for the placement process.

Ajay Shiv Sharma et al. [12] develops a predictor system for placement of student using logistic regression. The score in matriculation, senior secondary, subject in various semester of technical education & demographic of GNDEC's student is taken as input by the system. For the generation of result programming tool GNU octave & for optimization of classifier gradient descent algorithm, is used. It was revealed that 91% of total placement is from urban background, where gender played an important role as 59.09% of females are placed.

Ajay Kumar pal et al. [13] collected data of 65 students from MCA and evaluated using classification algorithm like Naïve Bayesian, C4.5 tree, and multilayer perception; for the prediction of training and placement. The attribute chosen from the database for evaluation include sex, MCA result, seminar performance, lab work, communication skill, and graduation background. Attribute assessment is done using chi-square test, information-gain and gain-ratio test. Average of these assessments is taken for each attribute and it was observed that sex has most impact on the output. Naïve Bayes classifier has highest accuracy rate of 86.15% with 0 time to build and lowest error 0.28.

Ramanathan.L et al. [14] uses EDM to predict the placement of final year student by using traits such as gender, category, academic gap, 10th -12th result, grade in B-Tech, communication skill, technical skill and grade in M-Tech. Similarity measure "Sum of Difference" is implemented on C# to analyze the pattern of data set and carry out prediction process. As a result of experiment a graph is obtained which shows that if value on y-axis goes higher than 4, it designate "yes" value of placement.

V.Ramesh et al. [15] scrutinize various data mining technique in order to use them for the purpose of placement prediction. A dataset of 300 students from computer science is taken and cleaned, then suitable attribute like English marks, Math marks, programming language, lab marks are chosen on which technique like Naïve bayes, SMO, j48, multilayer perception, Rep tree algorithms are applied. Data set is applied on Weka tool and it is observed that multilayer perception shows an accuracy rate of 87.395%.

Vikas Chirumamilla et al. [16] uses CGPA, backlog, aptitude, technical articulate core skill level & achievements as evaluation attributes, which are preprocessed and cleaned before applying naïve bayes and C4.5. For the analysis rapid miner tool is used. Using confusion matrix it is observed that Naïve bayes has accuracy of 77.78% and C4.5 has 88.89 %.

Ravina Sangha et al. [17] aims to develop a model which can extract information from the database of educational institution. Rule based classification using fuzzy is applied on the dataset of MESCOE pune. There are several advantages of using fuzzy as it gives various prospects rather than a single value, with its high efficiency in handling erroneous input.

G.Vadivu et al. [19] uses a database of educational institute containing large amount of student data which can be used to forecast student performance in forthcoming semester and likelihood to get placed in coming recruitment session. For this purpose two algorithm KNN and naïve bayes are used. Dataset of 250 students with 59 attribute is taken for experiment purpose & accuracy of 95.33 was observed in case of KNN & 97.67% in case of Naïve bayes.

Dammalapati Rama Krishna et al. [20] aimed to pinpoint whether the student will be placed or not. In order to improve their performance, improved decision tree algorithm is used. Attribute include city, communication skill, technical skill, grade, attitude, economical background, written test. Paper also includes privacy preserving approach. Improved C4.5 decision tree based algorithm is applied on the eligible students as it can handle both missing data and continuous data.

Clustering Approach: - Similar to classification it sorts data into classes; however unlike classification here the labels of class are not defined; therefore known as unsupervised classification. Clustering is based on maximum resemblance among objects in same class and minimum among different class. It is not specific to single algorithm and can be achieved through many algorithms depending upon the scenario where training set is not needed. Clustering is also used for exploring data as it finds natural grouping and because of this reason it is a useful data preprocessing step as it identify homogeneous groups.

Dr Rajan vohra et al. [11] aims to find the eligible candidate for the placement by segregating the students into three clusters i.e. group 1 contain student who will face difficulty in completing their degree, there score range is 0.55-0.6, count of such student is 170. Group 2 contain above average student they are eligible for placement, score range 0.7-0.8; count of such student is 230. Last group contain 200 students with a score range of 0.625-0.675 this group need to improve a bit to become eligible. The data related to academic detail, technical skill, project detail, training and soft skill act as input for weka tool where K-mean clustering & j-48 algorithm are applied. Here classification algorithm is used to predict the cluster for pre-final year student which results in some errors.

Karan Pruthi et al. [18] uses data mining techniques to extract information of student from previously recorded data, which can be used for prediction purpose. The paper not only tells about placement but also whether the student gets placed in Core Company or consultancy, as well as predict company's name. After the experiment it was observed that clustering, classification and naïve bayes has accuracy of 95.52%, 82.39%, 62.3% respectively while predicting company type , also classification and naïve bayes has accuracy of 62.1%, 45.78% respectively in case of predicting company's name.

3. A Comparative Study of Papers

Sr No.	Papers name (year)	Method/Algorithm	Best Method/Method : Accuracy
1	Paper 1 (2011)	1.Decision Tree 2.Naive bayes 3.Artificial neural network	Decision Tree
2	Paper 2 (2016)	1.Bayesian method 2.Multilayer perceptrons 3.Sequential minimal optimization 4.Decision Tree	Decision Tree (j48)

3	Paper 3 (2014)	1.ID3 2.CHAID 3.C4.5	ID3: 95.33%
4	Paper 4 (2017)	1.Classification	
5	Paper 5 (2015)	1.Bayesian theorem	
6	Paper 6 (2016)	1.Fuzzy logic 2.K nearest neighbor	Fuzzy logic: 92.67% KNN : 97.33%
7	Paper 7 (2017)	Classification(C5.0)	C5.0 : 75%
8	Paper8 (2015)	Decision Tree	Decision Tree(ID3):95%
9	Paper 9 (2014)	1.J48 2.Simple cart 3.Bayes net 4Rep tree algorithm 5.Naive bayes	Naïve bayes Bayes net
10	Paper 10	Fuzzy inference system	
11	Paper 11 (2015)	1.K-mean 2.clustering J-48	
12	Paper 12 (2014)	1.GNU octave 2.Gradient descent algorithm	
13	Paper 13 (2013)	1.Naive Bayesian 2.C4.5 3.Multilayer perception	Naïve Bayesian:86.15%
14	Paper 14 (2014)	Sum of difference	
15	Paper 15 (2011)	1.Naive Bayes 2.SMO 3.j48 4.Multilayer perception 5.Rep tree algorithm	Multilayer perception:87.395%
16	Paper 16 (2014)	1.C4.5 2.Naive Bayes	Naïve bayes: 77.78% C4.5: 88.89%
17	Paper 17	Rule based classification using fuzzy	

	(2016)		
18	Paper 18 (2015)	1.Clustering 2.Classification 3.Naive bayes	Company type: Company name Clustering :95.52%: Clustering: 55% Classification :82.93%: Classification: 62.1% Naive bayes:62.3%: Naïve Bayes :45.78%
19	Paper 19 (2017)	1.K nearest neighbor 2.Naive bayes	KNN: 95.33% Naïve bayes: 97.67%
20	Paper 20 (2014)	Improved C4.5 decision tree	

- Techniques used in papers

SR Number	Technique	Paper used
1	Decision Tree	Paper 1,2,8,20
2	Naïve Bayes	Paper 1
3	Artificial Network	Paper 1, 5, 15, 16, 18, 19
4	Bayesian Method	Paper 2, 5, 13
5	Multilayer Perception	Paper 2, 13
6	Sequential min optimal	Paper 2
7	ID3	Paper 3
8	CHAID	Paper 3
9	C4.5	Paper 3, 13, 16, 20
10	Fuzzy Logic	Paper 17, 10, 6
11	K nearest neighbor	Paper 6
12	C5.0	Paper 7
13	J48	Paper 9, 11
14	Rep Tree Algorithm	Paper 9
15	K mean	Paper 11
16	Gradient Descent Algorithm	Paper 12
17	Sum of difference	Paper 14
18	Clustering	Paper 20

Table -1: Sample Table format

Preparation of Manuscript			
Margins : Top	0.5"	Bottom	0.5"
Left	0.5"	Right	0.5"
Margin : Narrow	Font	Cambria / 10 pt	
Title of paper : 16 Point	Heading	13 Point	
Sub Heading :12 Point	Spacing	Single line spacing	

4. CONCLUSION

Various classification and clustering techniques are inspected to evaluate the performance of students in the recruitment procedure. Using the comparative study amongst these techniques ID3 with accuracy 95.33%, KNN with 97.33%, C4.5 with 88.89%, Naïve bayes with 86.15 %, Multilayer perception having 87.395% accuracy is suggested to be the best one. If placement cell conduct workshop in advance for pre-final year to train them according to the result of these techniques, students will be able to identify their scope of improvement and will be able to refine themselves accordingly.

5. Future Scope

It would of great help if we revise and update our curriculum and other extra activities for each semester in accordance with the public, private and government sector requirement. We can also predict which company picks which category of students. Make a list of skill a particular company looking for, then on the basis of that we can train our student. These traits will make prediction process more accurate.

REFERENCES

1. Sudheep Elayidom, Summan Mary Idikkula, Joseph Alexander, " A generalized data mining framework for placement chance prediction problems" International journel of computer applications(0975-8887) volume 31- No.3,October 2011.
2. Tripti Mishra, Dharminder Kumar, Sangeeta Gupta, "Students' employability prediction model through data mining" International journal of applied engineering research ISS0973-4562 volume 11, November 4 2016.
3. T.jeevalatha, N.Ananthi, D.Saravana, "Performance analysis of undergraduate students placement selection using decision tree algorithms" International journal of computer application (0975-8887) volume 109-No.15, December 2014.
4. Neelam Naik, Seema Purohit, "Prediction of final result and placement of students using classification algorithm" International journal of computer application (0975-8887) volume 56-No.12, October 2017.
5. Pratiyush Guleria, Manu Sood, "Predicting student placement using Bayesian classification" Third international conference on image processing 978-5090-0148, 2015 IEEE.
6. Mangasuli Sheetal B, Prof. Savita Bakare, "Prediction of campus placement using data mining algorithm-fuzzy logic and K nearest neighbor" International Journal of advance research in computer and communication engineering (2278-1021) vol-5, June 2016.
7. Revathy S, Roopika G, Rishitha R, Revathy P, "An approach to suggest company specific placement opportunities using data mining techniques" IJCSMC (2320-088X) vol-6, March 2017.
8. Namita Puri, Deepali Khot, Pratiksha Shinde, Kishori Bhoite, Prof Deepali Maste "Student placement prediction using ID3 algorithm" IJRASET (2321-9653) vol-3,March 2015.
9. Mosima Anna Masethe, Hlaudi Daniel Masethe, "Prediction of work integrated learning placement using data mining algorithm" WCECS volume 1 22-24 October 2014.
10. Ravi Kumar Rathore, J.Jayanth, "Student prediction system for placement training using fuzzy inference system" ICTAT journal on soft computing (2229-6959) volume 07

11. Praveen Ram, Dr. Rajan Vohra, "Generating placement intelligence in higher education using data mining" International journal of computer science and information technologies (0975-9646) volume 6 2015.

12. Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar, "PPS- placement prediction system using logistic regression" International conference on MOOC, Innovation and technology in education (978-4799-6876) IEEE 2014.

13. Ajay Kumar Pal, Saurabh Pal, "Classification model of prediction for placement of students" Modern education and computer science (49-56) November 2013.

14. Ramanathan.L, Swamalatha P, D. Ganesh Gopal, " Mining educational data for students' placement prediction using sum of difference method" International journal of computer application (0975-8887) volume 99 18, August 2014.

15. V. Ramesh, P.Parkavi, P.Yosodha, "Performance analysis of data mining techniques for placement chance prediction" International journal of scientific & engineering research (2229-5518) volume 2 , 8 August 2011.

16. Vikas Chirumamilla, Bhagya Sruthi T, Sasidhar Velpula, Indira Sunkara, "A novel approach to predict student placement chance with decision tree induction" International journal of system and Technologies (ISSN0974-2107) volume 7 , Issue 1,2014, pp 78-88.

17. Ravina Sangha, Akshay Satras, Lisha Swamy, Gopal Deshmukh, "Student's placement eligibility prediction using fuzzy approach" International journal of engineering and techniques(ISSN:2395-1303)-volume 2, November-December 2016 pp 148-152.

18. Getaneh Berie Tarekegan, Dr. Vuda Sreenivasarao "Application of data mining in predicting placement of students" International Journel of Research Studies in Computer Science and Engineering (IJRSCE) (ISSN 2349-4840(print) & ISSN 2349-4859(Online)) Volume 3, Issue 2, 2016, pp 10-14

19. G.Vadivu, K.sornalakshmi, "Applying machine learning algorithms for student employability prediction using R" International journal of pharmaceutical sciences review and research (ISN 0976-044X) March-April 2017, Article No. 11, pages 38-41.

20. Krishna, Dammalapati Rama, Bode Prasad, and Teki Satyanarayana Murthy. "Placement Prediction Analysis in University Using Improved Decision Tree Based Algorithm."