A Review on Student Placement Chance Prediction

Liya Claire Joy
Student
Department of Computer Science and Engineering
Rajagiri School of Engineering and Technology
*Ernakula*m
liyaclaire01@gmail.com.

Asha Raj
Assistant Professor
Department of Computer Science and Engineering
Rajagiri School of Engineering and Technology
Ernakulam

***Abstract*— All students dream to obtain a job offer in their hands before they leave their college. A placement chance predictor helps students to have an idea about where they stand and what to be done to obtain a good placement. A placement predictor is a system that could predict the possibility or the type of company a pre-final year student have chances to be placed. Thus a prediction system could help in the academic planning of an institution for future years. With the emergence of data mining and machine learning, many predictor models were introduced by analyzing the previous year student's dataset. This paper presents a literature survey on different placement prediction models for pre-final year engineering graduate students.**

Keywords*— Prediction, Data mining, Logistic Regression, Decision tree, Random Forest.

I. INTRODUCTION

According to statistics 1.5 million engineers are graduating each year in India. The demand and need for qualified graduates in field of IT industry is rising day by day. But most of the students are unaware about the needs of the IT industry. The number of the student graduates who satisfies the requirements and quality of a company is very low. Placements are one of the biggest challenge faced by a student in the lifetime. It is the responsibility of the institutions to provide maximum placement chance to its students. Also the placement cell and teachers of an institute should take proper steps inorder to produce a set of students suitable for each company's requirements. A placement prediction system can be used to identify the capability of a particular student for the specified job.

All companies in the IT sector spends a large amount of its total capital in recruiting the students to its company. Thus it is necessary to find an alternative process of filtering to reduce the capital cost that is used for this process. Effective filtering of students could be performed by applying various data mining and machine learning tools on the student details. Luan [1] defined the meaning of data mining in the field of education as a method of identifying, discovering and capturing the unknown similarities or patterns from a dataset by using an ensemble combination of various analytical approaches. It is possible for an educational institute to exploit this data mining feature to figure out the recruitment policy of a company from previous year placement statistics and student dataset. So the placement cell of the institution could prepare a placement predicted list for the present students. Thus it is very important to conduct a study on various placement prediction systems. This paper presents a survey on different placement prediction system models and its application for the students.

II. LITERATURE SURVEY

A. Prediction using Logistic Regression

This paper [2] presents the design of a placement predictor using the predictive analysis model called as Logistic Regression. Logistic regression is one of the most commonly used statistical model which is used as a classifier in the field of machine learning. The tool designed here predicts the probability of a student being placed and classifies the dataset based on prospect of getting recruited into a company or not. The dataset for the work consists of variables such as various marks obtained in secondary and graduation examinations along with demographic details such as resident status and gender of student. The dataset also comprises of a placement indicator variable to identify the placement status.

An optimization technique Gradient Descent Algorithm is applied on the training data to obtain the minimum values of the parameter that is used for classification. The minimization process is repeated until the decrease in the value of weight become negligible. The iterative step is given in (1)

$$\theta_j = \theta_j - \alpha \frac{1}{m} \sum_1^m \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)} \qquad (1)$$

Where,
α is the learning rate;
$x^{(i)}$ is any example of training parameter;
$h_\theta\left(x^{(i)}\right)$ is the hypothesis of the respective example;
$y^{(i)}$ is the actual outcome of the example;
m is the count of training examples.

The general formula to get the value of prediction for each parameter is given below in (2).

$$p = \theta^T x \qquad (2)$$

978-1-5386-9533-3/19/$31.00 ©2019 IEEE

Where,

p is the prediction value;

α is the corresponding optimized vector of weight to the parameter;

x is the vector value of the parameter.

Sigmoid function is applied on predictions inorder to obtain probabilities of classifier in the range of 0 and 1. The further tuning of classifier depends upon the difference between the actual and hypothesis value. As the subjects of the graduation exams are considered as the features in Eqn.1, the importance of the subject knowledge in the recruitment process is obtained. Thus the top skills focused by the company in their previous recruitment sessions could be obtained easily.

This paper [3] proposes a placement analyzer system that recommends students with the best suitable placement status depending upon their capabilities. The probability chances of students from different departments are predicted in this work. The five different placement statuses considered in this work are Dream Company (Companies with CTC ≥10 lpa), Core Company (Companies with CTC ≥4.5 lpa & CTC < 10 lpa), Mass Recruiters (Companies with CTC < 4.5 lpa), Not Eligible and Not Interested in Placements. The prediction is performed using Logistic Regression by using the biglm package in R tool. The dataset for the work includes basic details of student (such as gender, location), marks obtained and board of study in secondary examinations, graduation examination details (such as department, grade points and arrear history). The minimal value of each variable is computed using the regression analysis of different data for the variables found in the dataset.

The required probability chance of the system is computed using the (3).

$$P(Y) = e^{L}/(1 + e^{L}) \quad (3)$$

Where L is the aggregate logit sum of the minimal variables in the dataset.

B. Prediction using Fuzzy Approach

This paper [4] proposes a system to predict the eligibility and the improvement action levels required by students in the campus placement using a combination of Fuzzy approach and Rule based classification method. The ensemble of two approaches is applied on the student dataset containing academic and placement details to perform the prediction. Initially the values of attributes are defined in terms of linguistic variables, e.g., an attribute “eligibility” could take linguistic values such as {High, low etc.} indicating the possible chances of the students eligibility. A set of rules are built from the dataset using Rule based classification. All the rules in the knowledge base are evaluated to identify the necessary improvements required to become eligible for placement for the required students.

C. Prediction using Decision Tree Algorithm

This paper [5] proposes a model that predicts the probability of placement of a student in a company using ID3 decision tree algorithm. This system analyses the given dataset to identify the most relevant parameters required for placement prediction from the student dataset. Entropy and Information gain values of all parameters in the dataset is measured and the parameter with suitable measurement value is selected as split variable while building the decision tree. The Weka Tool generates an optimized decision tree with leaves representing the placement prediction chance of the student. The dataset comprises of marks obtained in secondary examinations, graduation grade points, arrear history and department type, details of various skills such as programming skill and communication skill, internships attended and details regarding interests in future studies.

Let the selected parameter has c different values and Pi be the associated probability value for each respective parameter, then the formulae for entropy measurement of each parameter is given in (4).

$$Entropy\ (s) = \sum - Pi\ log_2 Pi \quad (4)$$

The equation for information gain is given in (5) as the difference between entropy of original dataset and entropy of the subdivided dataset after selecting the spilt attribute.

$$Gain = H(D) - \sum P(Di)\ H(Di) \quad (5)$$

This paper [6] proposes a system to predict the possibilities of student placement selection using various decision tree algorithms. The most common decision tree algorithms such as ID3, CHAID, C4.5 and CART algorithms were applied on the dataset using the Rapid Miner Tool. The analysis is to figure out the most suitable algorithm for the given dataset. From the result analysis and measurements they found ID3 algorithm as the one with highest accuracy.

D. Prediction using Random Forest Algorithm

This paper [7] proposes a method for predicting the employability status of the student using Random forest algorithm. The dataset for the work consists of scores collected from the students by conducting a test for them in the areas concentrated for recruitment process. Random Forest is an ensemble prediction method by aggregating the outcomes of the individual decision trees [8]. The accuracy of the model trained using this algorithm can be improved by tuning the algorithm parameters such as number of trees and number of the attributes that is selected randomly. A model is created using Random forest that can predict the likelihood of a student to be placed in a company. From the trained model the system can display the name of companies a student have chances to be placed based on their obtained scores. The system can also display a list of company seeking skills to be incurred by the students who are attending the placement process. From the analysis performed in this work using different algorithms (such as

SVM, KNN, Decision tree, Multi-class Ada), they concluded that the accuracy produced by Random Forest is the highest for the given dataset.

E. Prediction using classification and clustering techniques

This paper [9] proposes a system that predicts the type of the company such as Consultancy or IT Company and the specific name of the company a student have chances to be placed based on their academic performance. The dataset comprises academic details of students including their grade points and performance details of the selected subjects as well as the recruited company details. Classification and clustering techniques are implemented using the J48 decision tree algorithms using the WEKA data mining tool. J48 is an extension of ID3 algorithm with some added features. A Naïve Bayes classifier model is also implemented using the WEKA tool. This supervised learning classifier is a statistical method based on the Bayesian theorem. This classifier is mostly used when the dataset is small with large dimensionality. Equation (6) states the Bayesian Theorem [10].

$$P(A/B) = P(B/A) * P(A)/P(B) \quad (6)$$

Where A is the hypothesis to be tested and B is the evidence associated with A. From the result analysis process, the system could identify the most featured attributes of the dataset in the recruitment process.

F. Prediction using Sum of Difference method

This paper [11] predicts the likelihood of placement obtained by a student in a recruitment process by using a similarity measurement model. In this work the mathematical model is built by using the similarity measure called Sum of Difference (SOD) method. The similarity measure is a procedure that is performed in order to obtain a pattern from the given dataset. The dataset is collected by conducting a survey among students and obtaining their details. Dataset comprises of personal details such as gender and category; academic details including grade obtained in various such as 10^{th}, 12^{th}, graduation and post-graduation exams, arrear history and gaps in between academic life; communication skills; details regarding the attended technical courses and placement status. A priority value is set for each attribute in the dataset and a combination of most required attributes are selected for prediction. Using the sum of difference method, a reference value is computed corresponding to the selected attributes. If a student scores above this value indicates that student will get placed in the recruitment.

G. Prediction using Job Competency Modeling

This paper [12] proposes a system that builds a Job competency model which consists of two phases. Initially all the domain fields required for each specific job are recognized and then a job competency score is calculated for students by analyzing their academic score in the domains recognized. If the computed score is greater than a threshold score indicates the student suites for the job designation. The input to the system includes the job designation, areas and course title required for job as well as the students marks obtained in graduate curriculum examinations.

For each student, a competency profile is formed as a graph by assessing the student's academic results. The root node of the graph is the job title followed by domains related to the job as the succeeding layers of the graph. The edges in the graph represents the relevance factor of each domain corresponding to the job and then the domain score is computed. Suppose node P is the starting node with n nodes, named as Qi, connected directly to the start node with weight Wi. Each node among n nodes is associated with a score Si and total score of the succeeding nodes at node Qi is Ti. The total score obtained by the student for the specific job designation can be obtained using (7).

$$S = \sum_{i=1}^{n} \left(\frac{Wi}{100}\right) * \frac{Si}{Ti} \quad (7)$$

III. CONCLUSION

A detailed study was conducted based on different placement prediction models. From the study it is clear that the student dataset containing academic and placement details are a potential source for predicting the future placement chances. This prediction can enlighten students to identify their capabilities and improve accordingly. This system also helps in the academic planning of an institution to prepare proper strategies and improve the placement statistics for the future years.

REFERENCES

[1] J. Luan, "Data mining and its applications in higher education", New Dir. Inst. Res, 113:17–36, 2002.

[2] A.S. Sharma, S. Prince, S. Kapoor, K. Kumar, "PPS – Placement prediction system using logistic regression", IEEE international conference on MOOC, innovation and technology in education (MITE), pp 337-341,2014.

[3] Thangavel, S.Bkaratki, P. Sankar, "Student placement analyzer: A recommendation system using machine learning", Advances in Computing and Communication Systems (ICACCS-2017) International Conference on. IEEE, 2017.

[4] R. Sangha, A. Satras, L. Swamy, G. Deshmukh, "Students Placement Eligibility Prediction using Fuzzy Approach", International Journal of Engineering and Techniques , Volume 2, Issue 6, Dec 2016.

[5] H. Bhatt, S. Mehta, L. R. D'mello, "Use of ID3 Decision Tree Algorithm for Placement

Prediction", International Journal of Computer Science and Information Technologies (IJCSIT), vol. 6, pp. 4785-4789, 2015.

[6] T. Jeevalatha, N. Ananthi, D. Saravana Kumar, "Performance analysis of undergraduate students placement selection using Decision Tree Algorithms", International Journal of Computer Applications, vol. 108, pp. 0975-8887, December 2014.

[7] Bharambe, Yogesh, "Assessing employability of students using data mining techniques", Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on. IEEE, 2017.

[8] P. Khongchai, P. Songmuang, "Random Forest for Salary Prediction System to Improve Students Motivation", 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 637-642, 2016.

[9] Pruthi, P. Bhatia, "Application of Data Mining in Predicting Placement of Students", International Conference on Green Computing and Internet of Things (ICGCIoT), 2015.

[10] P. Guleria, M. Sood, "Predicting Student Placements Using Bayesian Classification", International conference on Image Information Processing, IEEE Computer Society, 2015, pp. 109-112.

[11] 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.

[12] Soumya M.D, T. Sugathan, K. Bijlani, "Improve Student Placement using Job Competency Modeling and Personalized Feedback", Advances in Computing, Communications and Informtics (ICACCI), International Conference on. IEEE, pp: 1751 – 1755, December2017.