IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 5, May 2021

DOI 10.17148/IJIREEICE.2021.9507

Student Placement Prediction Using Support Vector machine Algorithm

Dr. Rajiv Suresh Kumar[1], Fathima Dilsha[2], Shilpa A N[3], Sumayya A A[4]

Department of Computer Science and Engineering, JCT College of Engineering and Technology,
Coimbatore, TamilNadu, India[1-4]

Abstract: Campus placement plays a vital role in every educational institution in helping students to achieve their goals. All students dream to obtain a job offer in their hands before they leave their college. In this paper, a predictive model is designed which can predict whether a student get placed or not. The main objective of this project is to analyze the student's academics data, aptitude data and predict the placement possibilities of students to have an idea about where they stand and what to be done to obtain a good placement. Which also aids to increase the placement percentage of the institutions? The data has been collected by the institution for which prediction is going to be done and by applying suitable data pre-processing techniques. The model is built by both training and test set which gives accuracy in prediction. Here we use a single supervised machine learning algorithm named support vector machine algorithm. This algorithm independently predicts the results and we then compare the efficiency of the algorithm, which is based on the dataset. This model will help the placement cell to focus on the potential students and help them to improve their technical and other skills.

Keywords: Prediction, SVM, 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.

II. RELATED WORK

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)

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

DOI 10.17148/IJIREEICE.2021.9507

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

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

Sigmoid function is applied on predictions inorder to obtain probabilities of classifier in the range of 0 and 1. 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).

$$P(Y) = e^L/(1 + e^L)$$

(3)

B. 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 be the associated probability value for each respective parameter, then the formulae for entropy measurement of each parameter is given in (4).

$Entropy\ (s) = \sum - Pi \log_2 Pi$(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.

$Gain = H(D) - \sum P(Di)\, H(Di)$(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.

C. 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].

$$P(A/B) = P(B/A) * P(A)/P(B)$$

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.

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 10th, 12th, 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.

IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 9, Issue 5, May 2021

DOI 10.17148/IJIREEICE.2021.9507

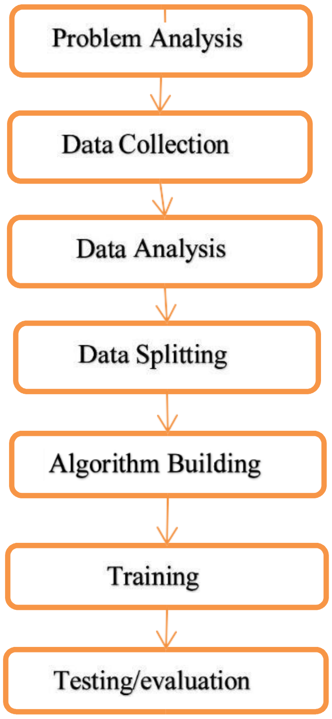
D. Prediction using Job Competency Modeling

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).

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

III. METHODOLOGY

In this paper we use machine learning techniques to predict the placement status of students based on a dataset. The parameters in the dataset which are considered for the prediction are Quantitative scores, Logical Reasoning scores, Verbal scores, Programming scores, CGPA, internal marks, external marks, list of students placed in a company The placement prediction is done by machine learning Algorithm using SVM.



1. Data Collection sample data has been collected from college placement department. As an input for model prediction, which consist of all the required dataset.

2. Data Preparation & Pre-processing

Data preparation is a step in a data analysis process in which data from one or more sources is cleaned, transformed and enriched to improve the quality of data prior to its use. The collected data were then pre-processed to fill the missing data and made compatible for further processing.

3. Data Splitting

Splitting the Dataset into Training set and Test Set ,Now the next step is to split our dataset into two. Training set and a Test set. We will train our machine learning models on our training set, i.e our machine learning models will try to understand any correlations in our training set and then we will test the models on our test set to examine how accurately it will predict. A general rule of the thumb is to assign 80% of the dataset to training set and therefore the remaining 20% to test set.

DOI 10.17148/IJIREEICE.2021.9507

4. Algorithm Building

SVM algorithm is appied on the dataset. SVM stands for Support Vector Machine. It is also a supervised machine learning algorithm that can be used for both classification and regression problems. However, it is mostly used for classification problems. A point in the n-dimensional space is a data item where the value of each feature is the value of a particular coordinate. Here, n is the number of features you have. After plotting the data item, we perform classification by finding the hyper-plane that differentiates the two classes very well. Now the problem lies in finding which hyper-plane to be chosen such that it is the right one. The Support Vector Machine (SVM) helps in identifying the hyperplane for classifying the data samples. In the case of multiple hyperplanes, the one which has maximum distance from the data points was chosen for better classification.

Advantages:

This algorithm performs best when there is a clear margin of separation.

- Effective in high dimensional spaces .
- If the number of dimensions is greater than the number of samples, the algorithm would be able to perform better
- It is memory efficient

5. **Evaluation and Testing:**

The performance measurement of the model was evaluated with the help of various metrics like accuracy, sensitivity, F1-score and precision. The performance visualization of the multi-class classification problem was analyzed using a graphical plot AUC (Area under the Curve) ROC (Receiver Operating Characteristics) curve that reveals the analytical ability of a binary classifier system as its discrimination threshold. The ROC curve is generated by plotting the true positive rate against false-positive rates at various threshold rates. The best algorithm based on the performance parameters was selected to predict the placement category of students. Based on the details provided by the students, the placement category could be predicted and the result would be displayed along with the suggestions for further improvement.

IV.CONCLUSION

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 and It is clear that SVM gives an accuracy of 100. 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.

V. FUTURE SCOPE

The future enhancements of the project are to focus on adding some more parameters to predict better organized placement status. We can also enhance the project by predicting some solutions or suggestions for the output generated by the system.

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 Salary Prediction System to Improve Students
Motivation", 12th International Conference on Signal Image Technology & Internet-Based Systems (SITIS), pp. 637-642, 2016.

[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 Placement2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS) 545 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 forSalary 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- Systems (ICACCS-2017) International Conference on 112. IEEE, 2017.