

# Campus Placement Predictive Analysis using Machine Learning

Nikhil Kumar

Department of Computer Science and Engineering  
Galgotias University  
Greater Noida, India  
[26nikhilkumar01@gmail.com](mailto:26nikhilkumar01@gmail.com)

Ajay Shanker Singh

Department of Computer Science and Engineering  
Galgotias University  
Greater Noida, India  
[drajay.cse@gmail.com](mailto:drajay.cse@gmail.com)

Thirunavukkarasu K.

Department of Computer Science and Engineering  
Galgotias University  
Greater Noida, India  
[thiru.k@galgotiasuniversity.edu.in](mailto:thiru.k@galgotiasuniversity.edu.in)

E. Rajesh

Department of Computer Science and Engineering  
Galgotias University  
Greater Noida, India  
[rajresearch16@gmail.com](mailto:rajresearch16@gmail.com)

**Abstract**—Each student dreams of having a work offer in their hands before leaving college. A selection probability indicator lets students get an sense of where they're standing and what to do to ensure a decent selection. A placement predictor is a device that can forecast the probability or form of business that a student in the pre-final year has chances of placing. While a forecasting program could help in the academic preparation of an institution for future years. With the emergence of data mining and machine learning, through analyzing the data set of the previous student year, numerous predictive models were applied. This paper introduces a literature study for pre-final year engineering graduate students on different statistical selection models.

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

## I. INTRODUCTION

According to statistics 1.6 million students pass from CS department every year. The demand for skilled and qualified students is rising day by day. Thus the company use a good amount of capital in recruiting students from in-campus and off-campus because number of skilled and qualified students are very low. Colleges and Institution needs to focus on practical knowledge of real world rather than completing their syllabus. Placements are the biggest opportunities in the life of a student and they need to be fully prepared while attempting it.

Placement Predictor system helps in predicting whether a student will get placement or not. This system can also be helpful for identifying the areas where student needs to work on for placement. This system uses student's details like academic marks, coding skills, etc. This system uses previous year placement statistics and student dataset for the placement

prediction so the placement cell of the organization could set up a placement anticipated rundown for the present students.

Along these lines it is important to direct an investigation on different placement prediction frameworks. This paper shows an overview on various placement prediction framework models and its application for the students.

## II. LITERATURE SURVEY

### A. Prediction using Logistic Regression

This paper[1] describes the creation of a indicator for placement utilizing the concept of predictive modeling, called Logistic Regression. This model is one of the most widely used mathematical model in the field of machine learning and is used as a classifier. The method developed here forecasts the probability of placing a student, and categorizes the dataset focused on the perspective of being or not being hired into a corporation.

The tool is designed by 4 steps:

- Process-

Logistic regression is a numerical methodology used to assess a preparation dataset that incorporates one or much more autonomous factors showed by  $X_b$  ( $b=0$  to  $N-1$ , for example  $N$  indicators) that construe an end for example Last figure. The outcome is estimated by methods for a dichotomous variable (in which there are just two potential outcomes). The needy variable in logistic regression is twofold or dichotomous, for example it contains just information coded as either 1 (TRUE, achievement and so forth) or 0 (FALSE,

disappointment and so on).  
 $\text{Logit}(p) = \pi = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots$

Where the chance of accomplishment is "p."

The logit progress is known as logging opportunity:

$$\text{Logit}(p) = \pi = \log(p/1-p)$$

- Model general fit-

The likelihood of invalid model - 2

Log is given by  $2 * \ln(L_0)$

Where  $L_0$  is the chance of causing estimations where autonomous factors to have little impact on the outcomes.

- Parameters of Regression -

Each factor or function is correlated with the  $\beta$  coefficient. This Coefficient measures the input in the Logistic System for increasing factor or independent variables. The further logistic model parameters add, the smoother the sigmoidal curve can produce. The Sigmoidal Curve scale is between 0 and 1.

- Final Processing-

At last, the data is applied on Logistic Model which was gathered from Online test of Students.

### B. Prediction using Fuzzy Approach

Fuzzy reasoning is essentially a philosophy that offers the direct understanding of reality and false values of which these attributes talk with the same guidelines here with the fuzzy logic, of which advice may be expressed of integrity and false esteems.

Compared to true or incorrect (1 or 0) Boolean logic, the principle of Fuzzy thinking is a way to deal with degree truth seeking. Fuzzy logic is an extension of multi-estimated propositional value logic to an enrollment degree array known as participation feature that maps into fuzzy sets or required selection of fuzzy sets known as fuzzy relationship.

The central aim of this paper[2] is to order broad selection of student data set utilizing fuzzy reasoning and to estimate readiness to position students on whether or not the applicant is qualified to apply for the placement. Two classes for a single undergraduate year, lessons for placement and non-placement preparation. These students will obtain better experiences or good CGPA in placement preparation training and the remaining student will go for additional lessons to improve their results. Both students can hardly be foreseen by this technique at any moment.

With the help of fuzzy logic, this approach essentially predicts and disintegrates parcels of student data collection for predefined classes which can be a good advantage for company to analyze data from vast volumes of student data which datasets.

The methodology in this paper[3] is elaborated using the methods of data mining. The equations used to create the model are "Fuzzy logic" and "K nearest neighbor" "Fuzzylogic is an approximate rather than exact logical framework for reasoning." The "K Nearest Neighbor (KNN) is a typical approximation of classification that combines, accessible case and groups the new cases based on the separation steps." Increasing model's efficiency / accuracy is visualized and checked, and each model results are addressed based on the show exam.

The algorithms are tailored to the collection of data and the characteristics required for product manufacture. The consistency gained after KNN analysis is 97.33 percent and 92.67 percent for Fuzzy logic. Subsequently, it would be easier from the preceding analysis and estimation if the KNN were used to forecast the outcome of the placement.

### C. Prediction using Random Forest Algorithm

This paper[4] uses the Random Forest Method to estimate employability status for the applicant. In a process-related analysis of hiring the data collection is collected from the candidates. Random Forest is a statistical technique for the ensemble by summing up the effects of increasing decision trees[5]. A forecast is generated using Random Forest, which may estimate the possibility of putting an understudy in an organization. The system will reveal the names of entities in the prepared model where an understudy has incentives to be placed based on the ratings they have received. Similarly, the program should have an summary of the business seeking skills that would be activated by the understudies that go through the situation process. We inferred from the research performed in this review utilizing unique formulas (e.g. SVM, KNN, Decision tree, Multi-class Ada) that the accuracy given by Random Forest is the most noteworthy for data collection.

### D. Prediction using Decision Tree Algorithm

In[6], author proposes to make a distinction key aspects that depend on the subjective and quantitative parts of a student's profile, such as CGPA, academic achievement, highly skilled and relational skills, and to structure a model that can predict a student's placement. It is for this purpose that ID3 classification methodology was used based on the tree of choosing.

With these attributes, a choice tree utilizing Weka tool is acquired. The J48 classification calculation which is

an augmentation of ID3 calculation is utilized to produce the choice tree.

The calculation gives a pruned choice tree with leaves as the choice that is set or not set. The essential node comprises of programming skills which can suit three potential qualities viz. Great, Average and Poor. On the off chance that the programming skills are Poor, the student isn't put. Moreover, if the programming skills are great, the student might be set dependent on the scholastic qualification which is CGPA. On the off chance that the CGPA of the student is over 7, student will be put generally the student won't be put. Additionally, if the student has normal programming skills, he may in any case be put dependent on different attributes like entry level positions, future examinations, relational abilities, and so on.

This paper[7] suggests a technique for predicting the possible effects of the campus placement assessment using various algorithms for the decision tree. For eg, CART, ID3, CHAID, C4.5 algorithms were added to the dataset by the Rapid Miner Method, the most well-known decision tree algorithms. The test is to take account of the most fitting equation for the specified dataset. They discovered ID3 measurement as the one with the greatest accuracy from the performance exam and estimations.

#### *E. Prediction using classification and clustering techniques*

This paper[8] suggests a hybrid solution utilizing two "K Implies Clustering and Support Vector Clustering" clustering algorithms, and one "Naïve Bayes" classification algorithm on the very same dataset.

The calculations used were performed, and the tool's front end was created using PHP and MYSQL as database.

Among all these calculations, K-Means ends up being the best predictive calculation speaking to bunch model, for taking care of predictive placing chance issues. The results were contrasted after using the calculation viz. ,K-Means (83 percent) and results were obtained with the help of Rapid Miner (83 percent). With the data generated by our investigation, students will now have the choice of choosing the correct specialization with the best chances of getting set.

#### *F. Prediction using Sum of Difference method*

Sum of Differential Strategy has been used in the paper[9] to achieve the prediction goal from the given dataset. The paper is managed using a proposed model using Sum of Differential Strategy to find most basic properties in a given dataset between things.

Apparently positioning on the grounds that it relies on various characteristics is not so straightforward to foresee, although the paper with four attributes is regarded. The paper selected the perfect mix of attributes. For a specified dataset, this blend works admirably. This is appropriate from the given dataset to determine the student's placement status. Therefore the model's performance is considered remarkable.

## II. CONCLUSION

A point-by - point analysis was performed based on specific statistical frameworks for the placement. The student dataset including academic and selection subtleties is clearly a possible hotspot from the exam for forecasting future selection possibilities. Such forecast will empower students to consider their skills and develop according to their needs. This methodology also helps to prepare appropriate procedures and strengthen the recruitment statistics for future years in the scholastic arrangement of an establishment.

## REFERENCES

- [1] Manoj K Shukla, Pranay Rambade, Jay Torasakar, Rakesh Prabhu, Prof. Deepali Maste, "Students Placement Prediction Model Using Logistic Regression", International Journal Of Engineering Research & Technology (IJERT), 2017.
- [2] Ravi Kumar Rathore and J. Jayanthi, "Student Prediction System For Placement Training Using Fuzzy Inference System", IJACT Journal On Soft Computing,2017
- [3] Mangasuli Sheetal, 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, 2016.
- [4] Bhamarbe, Yogesh, "Assessing employability of students using data mining techniques", Advances in Computing, Communications and Informatics (ICACCI), 2017.
- [5] P. Khongchai, P. Songmuang, "Random Forest for Salary Prediction System to Improve Students Motivation", Signal- Image Technology & Internet-Based Systems (SITIS), 2016.
- [6] Hitarthi Bhatt, Shraddha Mehta, Lynette R. D'mello, "Use of ID3 Decision Tree Algorithm for Placement Prediction", (IJCSIT) International Journal of Computer Science and Information Technologies, 2015.
- [7] Dammalapati Rama Krishna, Bode Prasad, Teki Satyanarayana Murthy, "Placement Prediction Analysis in University Using Improved Decision Tree Based Algorithm", International Journal of Computer Trends and Technology (IJCTT), 2014.
- [8] Syed Junaid Rahmathulla, "A Combined Clustering and Classification Approach for Predicting Placement Chance", International Journal of Engineering Research & Technology (IJERT), 2016.
- [9] Ramanathan.L, Swamalatha P, D. Ganesh Gopal, "Mining educational data for students' placement prediction using sum of difference method", International journal of computer application, 2014.