

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/372838094>

# Student Placement Analysis using Machine Learning

Conference Paper · June 2023

DOI: 10.1109/ICCES57224.2023.10192633

---

CITATIONS

10

READS

667

3 authors, including:



Rajalakshmi Raja

Sathyabama Institute of Science and Technology

29 PUBLICATIONS 111 CITATIONS

SEE PROFILE

# Student Placement Analysis using Machine Learning

Navuluri Divya<sup>[1]</sup>, Sravya Namburu<sup>[2]</sup>, Rajalakshmi Raja<sup>[3]</sup>

[1][2] UG Students, Dept. of CSE, Sathyabama Institute of Science and Technology, Chennai, India

[3] Assistant professor, Dept. of CSE, Sathyabama Institute of Science and Technology, Chennai, India

divya.navuluri347@gmail.com, sravyachowdary127@gmail.com, rajalakshmi.cse@sathyabama.ac.in

**ABSTRACT-** Every educational institution recognises the importance of campus placements in assisting the student in achieving their objectives. According to the criteria of the company, the analysis of the placements needs to create an estimate of the likelihood that students will be put in a particular organisation. The placement analysis uses a different number of parameters that can be used to evaluate the student's skill level. While some criteria are based on university standards, others come from assessments made using the placement management system itself. By combining these data points, the analysis must correctly forecast whether the student will be hired by a company. Finding a classification method that would work for our data collection and be as accurate as possible was the challenge. The type of problem an algorithm must solve and the data collection it must use will determine the accuracy of the algorithm. In order to analyse the accuracy levels of each algorithm regarding our problem and data set, this study ultimately choose SVM, Random Forest, Decision Tree as Existing algorithms and Logistic Regression, KNN, Gradient Boosting Classifier as Proposed Algorithms. This study may use the results of this test to decide which method to apply when integrating our analysis into the placement analysis system.

**Keywords-** Placement, Parameters, Accuracy, Naïve Bayes, Random Forest, Placement analysis, Logistic Regression.

## I.INTRODUCTION

Campus placement is regarded as the most important factor in India for determining how well a university or college is ranked and identified. Educational institutions are multiplying rapidly today. Every college or university's placement unit seeks placing students. Increasing student placement success is one of the biggest difficulties that higher education institutions confront today. As the calibre of instructional entities rises, the placement prediction become more challenging. One of the most effective techniques to quality improvement is to provide fresh knowledge about educational processes and entities into the managerial system. The data is usually retrieved using machine learning algorithms from operational and historical knowledge scattered across the databases of the academic institution. Information deployment includes information on

previous student data. These insights are used to assess the model's categorization abilities and to train the model to identify

rules. Placements have a significance to both students and educational institutions. It helps students build a strong foundation for their future professional jobs, and a strong placement rate gives university an edge in the competitive education market. This study focused on a system that determines a student's placement based on their credentials, historical information, and experience. A machine-learning system determines how this forecast will turn out. As part of creating a placement management system at the college level, we want to create a placement predictor that estimates the likelihood that students will be placed and aids in improving their abilities prior to the hiring process.

The student data from different universities for the year 2022 is the dataset taken into consideration for this project. We have taken the dataset of nearly three thousand members of parameters like gender, tenth grade, tenth board, intermediate grade, intermediate board, Engineering percentage, specialization etc. It includes various attributes like secondary and higher education percentage and board; it also consists of specialisation and work experience and many more. This study used machine learning algorithms to analyse students who were hired by companies during university placements based on a variety of factors.

For the placement analysis, machine learning is being used. To categorize students into the proper clusters, we take into account Existing methods include Support Vector Machine (SVM), Random Forest, Decision tree and K-nearest neighbour (KNN), Logistic Regression, Gradient Boosting Classifier as proposed algorithms. The results will assist the students to improve their profile. Respected algorithms' correctness is noted, and Recruiters and students would benefit from the comparison of different machine learning algorithms during placements.

## II.LITERATURE SURVEY

[1] Yang Liuet al. "The Application of Machine Learning Methods in College Students' Information System,". International Conference on Computer Science, Electronics, and Communication Engineering 2018 (CSECE 2018). Atlantis

Publications, 2018. The article appears education records application that seeks massive quantities of data from school administrators. The decision tree affiliation rules assessment was utilised for information pre-management. Understudies' scores were reduced by the use of usage, and reduced through the use of associations of the rule computations. It include confirmation technique was used to select the best component from the given dataset. As credits for the review, information assessment were employed.

[2] **Ishizue et al** Use of class attitude, mental scales, and code estimations as Markers of student condition and competence for programming classes. Studying Technology to Improve Learning 13.1, 2018. This research demonstrates how AI evaluations could be used to undermine student position assessments without exams by considering factors like mental acuity, programming tasks, and student attention to surveys. Using 9 illustrative components in a decision tree portrayal model resulted in a F-grade of 0.962. for However 20 predictable components using Support vector machine has standardised limited cumulated gain of 0.942 for the best-arranging model.

[3] **Ahmed S. et al.** "Execution Based Placement Prediction System".2018. The use of DM approaches in coaching is the main topic of this essay. A TPO board structure was setting up that could actually identify qualifying stowaways' road for the ground. The computation for the decision tree C4.5 was used to determine the affiliation's previous year data and current needs, which is crucial for understudies because the model would make a recommendation to qualified newbies, enabling them to determine whether they are prepared for it. They could then prepare in plenty of time for the grounds drive thanks to this. Academic history factors used for the review were rating marks, the range of the prowess, dominance, irrational factors, curiosity are all important factors.

[4] **Nichat, et al.** "Using Decision Tree Technique to Expect and Analyze Student Performance." Journal International 5, 2017. A model that might predict the understudy execution assessment using a decision tree classifier was organised by the author of this paper. The limit was evaluated using the demand technique while considering various factors associated to execution and understudy action. The results might anticipate the projected progress in the region and assist in making sound decisions to avoid disaster. This technique also includes an improvement area that the understudy must address to truly be satisfied.

[5] K-Nearest Neighbour's Classifier-Based Placement Prediction System. Animeshgiri, Vignesh V Bhagavat, Bysani Pruthvi, Nainidubey presented. The K-Nearest Neighbour's algorithm is used to forecast placement. classification to estimate the likelihood that students will be hired by different company. The outcome is also contrasted with outcomes from SVM and Logistic regression are two more machine learning models. Companies evaluate students' academic records as well

as their skill sets, including their ability to programme, communicate, analyse data, and operate in a team.

[6] **Goyal, J., et al.**, Positional Prediction of Decision Supporting System Using Data Mining, 4(2). With the aid of information mining analysis, the author of this study organised a position assumption that is extremely outstanding. The model developed assisted in determining situations probability and supported anticipating the number of the pupil was capable at handling social occasions. Improved Simple Bayes and Naive Bayes were used both taken into account. For information appraisal, WEKA and NetBeans instrument 6 were used. Results showed that when the 560models' dataset was taken into consideration for the review, As compared to Naive Bayes (81.96%), Figure Enhanced Naive Bayes produced an accuracy of 83.7%.

[8] **Sunitha R. et al.**, Int. J. Sci. Eng. Appl. Sci. 2.6 2016, Figure of Students Result Utilizing Machine Learning Approaches .Using the dataset from senior students, the author of this research created a data model for improved student grandstand relationships. J48 computation was employed, and various data mining analyses were done to isolate the model's level of precision. When placed apart Gradient boosting classifier and multi-layer regression perceptrons, for example 87.92% and 92.94% separately, J48 assessment reached high accuracy of 97.27%. The audit's chosen criteria included debts, grades, engineer cutoff, medium, and board type. In every manner, the execution documentation supplied was great.

### III.METHODOLOGY

The following figure illustrates the research approach followed in this work.

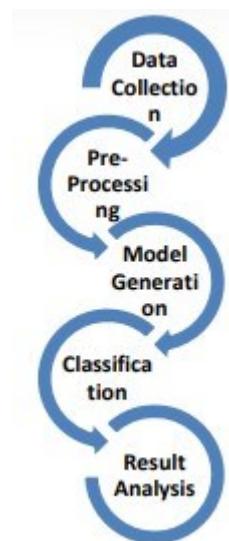


Fig. 1. METHODOLOGY

#### IV. EXISTING SYSTEMS

In India, 1.5 million engineers graduate each year, according to data. Graduates with the right qualifications are increasingly in demand in the IT sector. However, the majority of pupils are ignorant of what the IT sector requires. A very small percentage of student graduates meet the standards and standards of a corporation. One of the hardest obstacles a student will ever encounter is placement.

**1. Support Vector Machine (SVM):** Support The technical name is Vector Machine. Supervised Machine Learning algorithm method can be applied to classification and regression issues. However, classification issues are the main purpose for it. A data item that represents a point in n-dimensional space is one in which each feature's value corresponds to a specific coordinate. The number of features you have in this case is n. We accomplish classification after plotting the data item by locating the hyper-plane that effectively distinguishes the two groups. The challenge now is determining which hyper-plane should be picked for it to be the proper one, SVM can also be utilised using the Python module Scikit-learn, which can be used to create a variety of machine learning algorithms.

**2. K-Nearest Neighbours:** KNN is an abbreviation for K-nearest neighbours. This simple method may be used to solve classification and regression problems. Because it is a supervised machine learning approach, it employs labels. This algorithm's fundamental operation is based on the idea that related entities are always located close to one another. So, this is a presumption made in order for this algorithm to produce any useful results. KNN uses distance, closeness, or proximity to represent similarity. Distance is calculated mathematically, and the Euclidean Distance is typically used because it is the most.

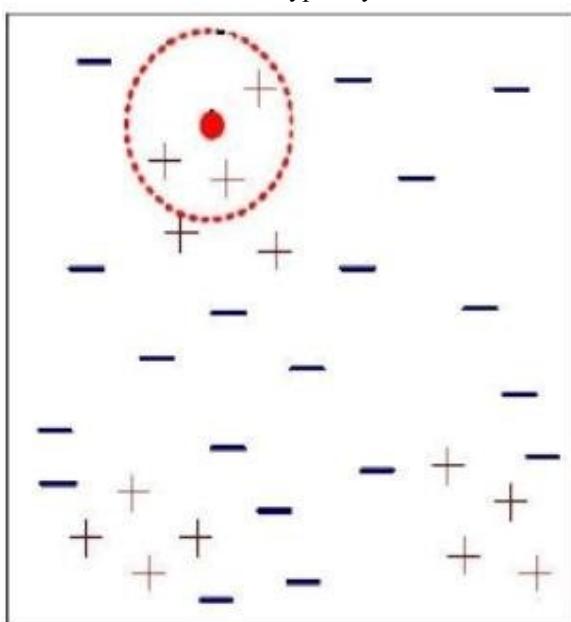


Fig.2. K-Nearest Neighbour

The algorithm is to load the necessary data.

Step 1: First, enter data set taken.

Step 2: In step two, apply KNN to the supplied parameters.

Step 3: Euclidean distance between the parameters now in place and computed for freshly inputted data taken in the dataset.

Step 4: Output is forecasted using similarity metrics.

**3. Decision Tree:** ID3 decision tree technique is used in this research to offer the models which predicts whether or not a student will be placed in a firm. Given dataset is analysed by this method to determine the student dataset's most pertinent parameters needed for placement prediction. All parameters in the datasets have their Entropy and Information Gain values tested, and while constructing the decision tree, parameters with good measurements values is choose as the best split variable. The weka tool generates optimal decision, with leaves representing the student preferences possibility of being placed where they are predicted to be. The dataset includes information about secondary test results, graduation grade points, past due history, department type, information about various talents including programming and communication, internships attended, and information about interests in future studies.

#### V. PROPOSED SYSTEMS

**1. Logistic Regression:** The logistic regression classification approach is ideal for binary values classification. Its decision is boundary and is often determined linearly from probability estimates. The results are in the form of a nonlinearly parameters form of estimation optimization problem. maximising the expression with any nonlinear optimization solver, parameters can be calculated. This method's objective is to identify the class from which a new data point is most likely to have come. Two quantitative and qualitative input qualities are possible. Rather than a hyperplane or straight line, logistic regression employs the logistic function to calculate the outcome of a linear equation between 0 and 1.

**2. Random Forest:** There are several classification methods at our disposal, including SVM, Logistic regression, Decision Tree, and Naive Bayes Classifier, to mention a few. The Random Forest Classifies, on the other hand, has a significant place in the classification hierarchy. We will investigate the function of decision tree since the random forest classifier is made up decision tree. Apart from the leaf node, each node indicates a test the feature, similar to a flowchart. The class label is represented by a leaf node, and the conjunctions of characteristics that result in those class labels are represented by branches. The pathways from the root node to the leaf node indicate a decision tree's classification rules. So, let us examine random forest classifiers right now. It is a collection of decision trees, as was previously described. "The wisdom of the multitudes" is the main tenet of random forest. It is a strong idea that several uncorrelated models, or in this case, trees, working

together would yield results that were considerably more reliable than those of the individual models. Therefore, in a random forest, each unique tree would look for the most acceptable class label for the issue using its unique attributes and classification criteria. Each tree would provide a unique response. The random forest conducts voting to determine which class label received the most support. The ultimate class classification for the issue would be whichever received the greatest support. This offers a model for class label prediction that is more precise.

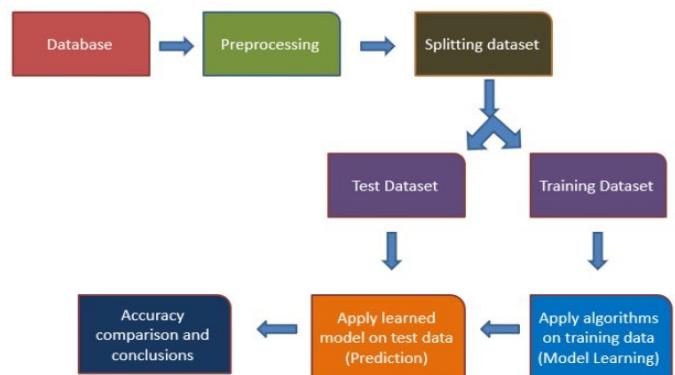
**3.Gradient Boosting Classifier:** Using the Python Scikit-Learn module, Gradient boosting is relatively straightforward. Before using Gradient boosting, the data set needs to be prepared. The data collection is then loaded into data frames. A multidimensional mathematical array is all that data frames are used to store each attribute's values. Since Gradient boosting is a supervised machine learning approach, label processing is challenging. In order to keep things simple, different encoding algorithms were used to encode the labels in data frames. similar to one hot encoding Processing on encoded data accelerates and simplifies the process. Categorical to numerical encoding is employed in this project. After that, we receive a data matrix that is prepared to be fed into ML models. To make our work easier, the data set is divided into two parts: training data and testing data. A common ratio is 70 and 30, which means 70% data will be utilised for training the model and 30% will be used for testing it.

By using means testing on the training data, the over-fitting issue can be avoided. A memorising system would result from using the same data for training and testing and hence it is crucial to keep training and testing data. Model evaluation, which compares the trained model to test data of unobserved dataset, is next step. Model's response is used to calculate its performance means how often it is accurate, how frequently it is unsuccessful, and how often it correctly predicts both true and false outcomes. On the basis of it, other criteria including accuracy, precision, and recalls were determined. The specifics of the performance evaluation criteria are discussed.

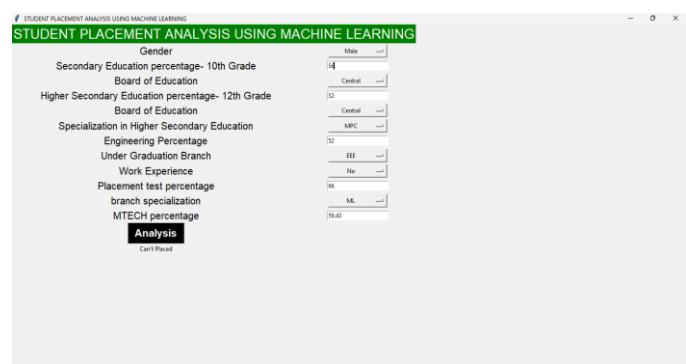
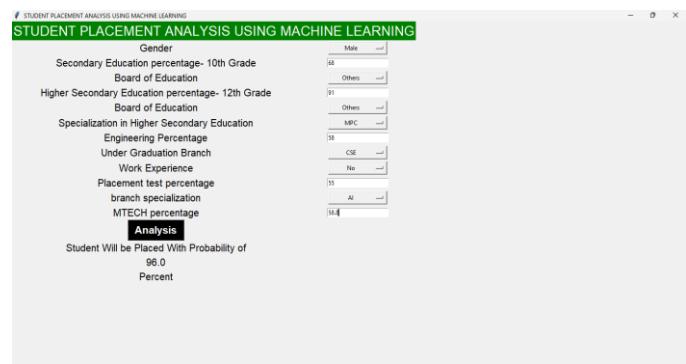
#### 4.GRAPHICAL USER INTERFACE:

In proposed system, we are including the Graphical User Interface for easily analysing the placement status of the students. Graphical user interface helps humans to easily interact with the electronic machines like computers. By using graphical user interface humans can easily express their ideas and intentions to perform electronically.

## VI. SYSTEM ARCHITECTURE



## VII.RESULT



## VIII.CONCLUSION

The likelihood of students being placed in campus drives increases as a result of the proposed method. Both students and the institution can benefit from the "STUDENTS PLACEMENT ANALYSIS USING MACHINE LEARNING". By being aware of the predictions made by this model, the school can concentrate on the possible students. The project primarily provides information regarding students' chances of being hired for campus drives, which is advantageous to both the students and organisation. At last we got best accuracy in proposed algorithms when compared to existing algorithms.

## IX.ACKNOWLEDGEMENT

We are really happy to publish our article on "Student Placement Analysis Using Machine Learning" and we would like to thank everyone who helped us complete this project by sharing their knowledge and assistance. We want to thank Dr. Rajalakshmi Raja, our project guide, for all the advice and inspiration she gave us when the project was being developed. The information in this paper is complete and pertinent. We have provided a good overview of the project we're building at the outset, and as we move forward, specifics about how the project will be implemented are discussed using technology.

## REFERENCES

- [1] Jositha Goyal and Shilpa Sharma, "Placement Prediction Support System using Data Mining", International Journal of Engineering and Techniques (IJET), (630-635), March-April 2018, ISSN:2395-1303.
- [2] Rajalakshmi Selvaraj and Sunitha Sivakumar, 2017.Adaptive Model for Improved Decision Tree Campus Placement Prediction. Engineering and Applied Sciences Journal, 12, 6069-6075. 18: 250-260 Asian Journal of Information Technology.
- [3] "A DATA MINING MODEL TO IMPROVE PLACEMENT," a study written by Ravi Tiwari and Awadhesh Kumar Sharma. Volume 120, Number 12, of the International Journal of Computer Applications was published in 2015.
- [4] D. Satish Kumar, S. Anusha, Zilina, Srinivasa. Rao "LOGISTIC REGRESSION PREDICTING STUDENT'S CAMPUS PLACEMENT PROBABILITY".
- [5] "Study on Students Placement Prediction Model Using Logistic Regression," International Journal of Research and Technology (IJRT).
- [6] "PS – Placement Prediction System using Logistic Regression" by Ajay Shiva Sharman, Swaraji Prince, Shubhan Kapoor, and Keshav Kumar Varma was published at 2014 IEEE-International conference on, Innovation, Technology in an Education.
- [7] "Study on Students Placement Prediction Model Using Logistic Regression," International Journal of Research and Technology (IJRT).
- [8] D. Satish Kumar, S. Anusha, Zilina, Srinivasa. Rao "LOGISTIC REGRESSION PREDICTING STUDENT'S CAMPUS PLACEMENT PROBABILITY".
- [9] "A DATA MINING MODEL TO IMPROVE PLACEMENT," a study written by Ravi Tiwari and Awadhesh Kumar Sharma. Volume 120, Number 12, of the International Journal of Computer Applications was published in 2015.
- [10] "PS – Placement Prediction System using Logistic Regression" by Ajay Shiva Sharman, Swaraji Prince, Shubhan Kapoor, and Keshav Kumar Varma was published at 2014 IEEE-International conference on, Innovation, Technology in an Education.
- [11] Rajalakshmi Selvaraj and Sunitha Sivakumar, 2017.Adaptive Model for Improved Decision Tree Campus Placement Prediction. Engineering and Applied Sciences Journal, 12, 6069-6075. 18: 250-260 Asian Journal of Information Technology.
- [12] "Study on Students Placement Prediction Model Using Logistic Regression," International Journal of Research and Technology (IJRT).
- [13] "PS – Placement Prediction System using Logistic Regression" by Ajay Shiva Sharman, Swaraji Prince, Shubhan Kapoor, and Keshav Kumar Varma was published at 2014 IEEE-International conference on, Innovation, Technology in an Education.
- [14] "Study on Students Placement Prediction Model Using Logistic Regression," International Journal of Research and Technology (IJRT).
- [15] Jositha Goyal and Shilpa Sharma, "Placement Prediction Support System using Data Mining", International Journal of Engineering and Techniques (IJET), (630-635), March-April 2018, ISSN:2395-1303.