

A CASE STUDY : ON INDIRA COLLEGE OF COMMERCE & SCIENCE STUDENT'S CAMPUS PLACEMENT DETERMINATION USING LOGISTIC REGRESSION ANALYSIS FOR PREDICTION

Prof. Shital S Pashankar

Computer Science Department, Indira College of Commerce and Science, Pune, India.

Email: srbhigwankar@gmail.com

Prof. Ashish Dhoke

BCA Department, Indira College of Commerce and Science, Pune, India.

Email: dhoke.ashish@gmail.com

Abstract

The placement of students is a critical objective of every educational institution. Institutions strive to strengthen their placement departments to improve their reputation and yearly admissions. A predictive model can be used to analyze previous year's student data and predict the placement chances of current students. This model can also suggest additional skills necessary for future recruitment, aiding students in their preparation for placement. The model can also be designed to calculate the possibility of a student being placed in a company based on the company's criterion. Parameter academic performance is used to assess the student's skill level. The placement predictor takes this parameter, combines them with data points from past students, and accurately predicts if the student will or will not be placed in a company. The proposed model can predict whether a student will be placed or not using a single supervised machine learning algorithm named support vector machine algorithm. This algorithm independently predicts the results, and the efficiency of the algorithm is compared based on the dataset. The model will help the placement cell to identify potential students and improve their technical and other skills, thereby increasing the placement percentage of the institution.

Keywords: ICCS, Machine Learning, Logistic Regression, Prediction

INTRODUCTION

The Outcome Based Education program is being adopted at planning institutions throughout India, and it revolves around student performance through monitored results, instilling skills, and observed attitudes. The establishment is given the choice of assessment processes for candidates during the program. In educational institutions, the student's position plays a significant role in uplifting institutional standards. Their academic performance and abilities are positively affected by placements. To achieve excellent positions, students should possess problem-solving skills, honesty, hard work, teamwork, and multitasking abilities. Therefore, a model is proposed that predicts the outcome of the student's conditions based on their previous academic performance.

The placement of students is crucial for the success of educational institutions, and it is determined by the campus placement of students. Therefore, a placement prediction system based on machine learning algorithms is being developed to predict the probability of a student being placed. The system considers the academic history of the student, such as their overall percentage, backlogs etc. It predicts the outcome of the student's conditions based on their previous academic performance. The goal is to develop a placement management system at the college level that helps students improve their skills before the recruitment process starts. The system uses Logistic Regression machine learning algorithm

to classify students into appropriate clusters, and the results help recruiters as well as students during placement-related activities.

The development of a placement predictor system based on machine learning techniques will help students put in more effort to make appropriate progress in various technical fields. The system can extract information from the educational organization's databases using operational and historical knowledge. The information set for the system implementation contains data on past students, which is used to train the model for rule identification and for testing the model for classification. The prediction of placement status that students are most likely to achieve will help students, teachers, and the placement cell in an institution provide proper care towards the improvement of students during the course.

Use the enter key to start a new paragraph. The appropriate spacing and indent are automatically applied.

Logistic Regression is one of the most used statistical models. It is used as a classifier in the field of machine learning [2]. It predicts the probability of a student being placed and classifies the dataset based on the 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. Gradient Descent Algorithm is applied on the training data to obtain the minimum values of the parameter used for classification.

LITERATURE SURVEY

Logical regression is a machine learning algorithm used for predictive analysis. This algorithm is widely used in various fields, including education, where it is used to predict student performance and placement. This literature review includes seven research papers that demonstrate the use of machine learning techniques to analyze and predict student information. In recent years, machine learning algorithms have gained significant popularity in various fields, including business, finance, healthcare, and education. One of the most widely used algorithms in predictive modeling is logistic regression. Logistic regression is a statistical method that is used to analyze a dataset and identify relationships between the dependent and independent variables. In this literature review, we will focus on a case study that uses logistic regression for prediction in the field of campus placement determination. Campus Placement Determination Using Logistic Regression Analysis for Prediction: The case study aims to determine the factors that influence the placement of students in campus recruitment drives. The study uses a dataset that includes information about the students' academic background, such as their percentage of marks in high school, higher secondary school, and undergraduate programs, as well as their specialization and work experience. The study uses logistic regression to predict the probability of a student getting placed based on their academic performance and other factors. The study found that the percentage of marks in the undergraduate program was the most significant predictor of placement, followed by work experience and the specialization of the student. The study also found that the probability of placement increased with an increase in the percentage of marks in the undergraduate program. For example, a student with 80% or above in the undergraduate program had a higher probability of placement compared to a student with less than 80% marks. The study concludes that logistic regression is an effective method for predicting the placement of students in campus recruitment drives. The study suggests that colleges and universities can use the insights gained from logistic regression analysis to improve the placement rates of their students.

METHODOLOGY

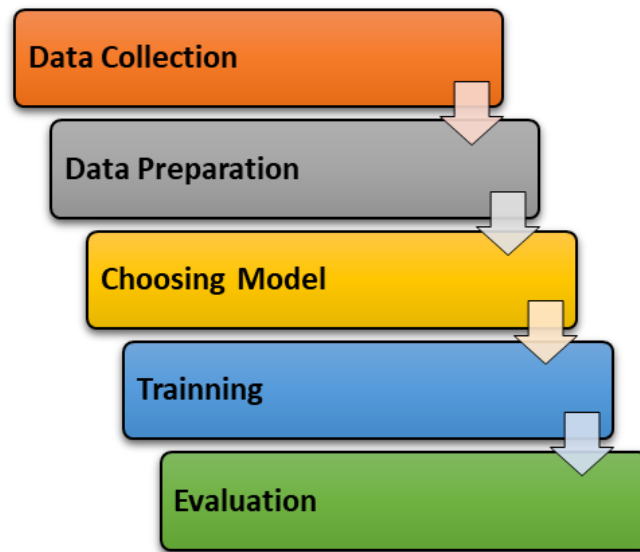


Figure 1 figure showing methodology used for research.

DATA COLLECTION

The sample data has been collected from Indira College of Commerce and Science (ICCS) college of BSc (Computer Science) and BCA third year students for academic year 2020-21. The dataset consists of 251 instances of students.

PRE-PROCESSING

The collected raw data sometimes has impurities or noise. So, data pre-processing method is used to convert raw data into clean data. This research work consists of pre-processing tasks such as attribute selection, finding and cleaning missing values. Once the data set is cleaned then splitting of the dataset into training and testing is done. Some attributes such as serial no. is removed as it does not contribute to classification.

CHOOSING MODEL

There are various machine learning algorithms available. We chose Logistic Regression for prediction of campus placement for given dataset. The proposed model will help in describing the data.

TRAINING

Now the next step is to split our dataset into two parts i.e Train set and Test Set. We will train our Logistic Regression models on our training set Now our model will try to understand the correlations present in our training set. After that we will test the logistic regression model on our test set to examine how accurately it will predict the result. For this research we have assigned 80% of the dataset for training set and the remaining 20% to dataset will be used as test set.

EVALUTION

The results obtained from different classifier is analyzed. The model was defined with the help of evaluations metrics like Precision, Recall, F1-Score and accuracy.

A. Data set for Logistic Regression

Table 1 shows the different parameters/ features taken into consideration to construct ICCS student data. The features include the student data from their class SSC percentage to Graduation, i.e., BSc

(Computer Science) and BCA percentage. Their Gender, SSC & HSC Board, Degree, Specialization, year down and backlog count.

Table 1. The data set comprises student data

Sr. No.	Features	Description
1	srno	This column defines the serial number of the student.
2	gender	This column defines the Gender of the student.
3	ssc_p	This column defines the 10th Grade percentage of marks of the student
4	ssc_b	This column defines the 10th Board of the student.
5	hsc_p	This column defines the 12th Grade percentage of marks of the student
6	hsc_b	This column defines the 12th Board of marks of the student
7	degree_n	This column defines the student's degree name in his/her graduation.
8	degree_p	This column defines the Degree Grade percentage of marks of the student
9	yd	This column defines the number of year down for the student.
10	bclg	This column defines the number of backlogs for the student.
11	spcl	This column defines the student's specialization in his/her graduation.
12	status	This column defines whether the student got placed or not.

	gender	ssc_p	ssc_b	hsc_p	hsc_b	degree_n	degree_p	yd	bclg	spcl	status
0	0	89.6	5	65.08	6	1	60.45	0.0	0	1	1
1	0	85.4	5	72.00	6	1	75.76	0.0	0	1	1
2	0	71.8	5	54.62	6	1	45.45	0.0	1	1	0
3	1	76.6	6	65.85	6	0	59.50	0.0	0	0	1
4	1	87.8	5	68.00	6	1	60.56	0.0	1	1	1
...
246	1	80.0	5	66.00	6	1	56.00	0.0	1	1	1
247	0	84.2	6	64.92	6	0	55.00	0.0	0	0	0
248	0	97.0	0	94.30	0	0	73.95	0.0	0	0	0
249	1	61.8	6	60.82	6	0	45.54	1.0	4	0	0
250	1	86.0	5	71.54	6	1	74.27	0.0	0	1	1

251 rows x 11 columns

Figure 2 showing preprocessed dataset.

Using the count plot, we tried to explain the frequency count of student placement data gender wise.

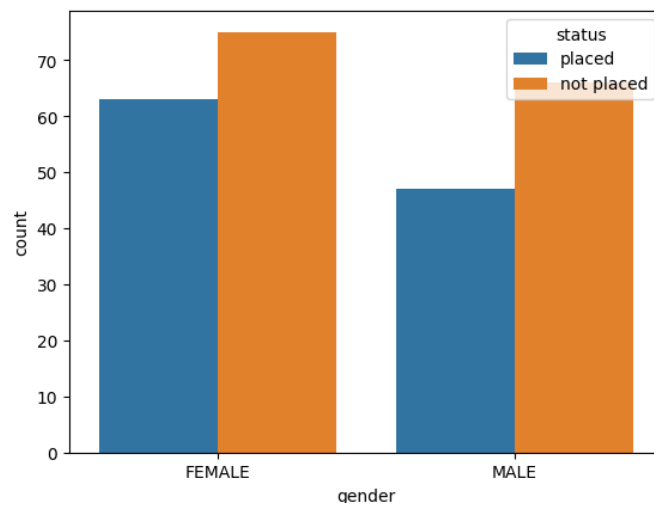


Figure 3 count plot checks the student's gender with the status of placement.

Using the pair plot method, we tried to explain the spread of student placement data.

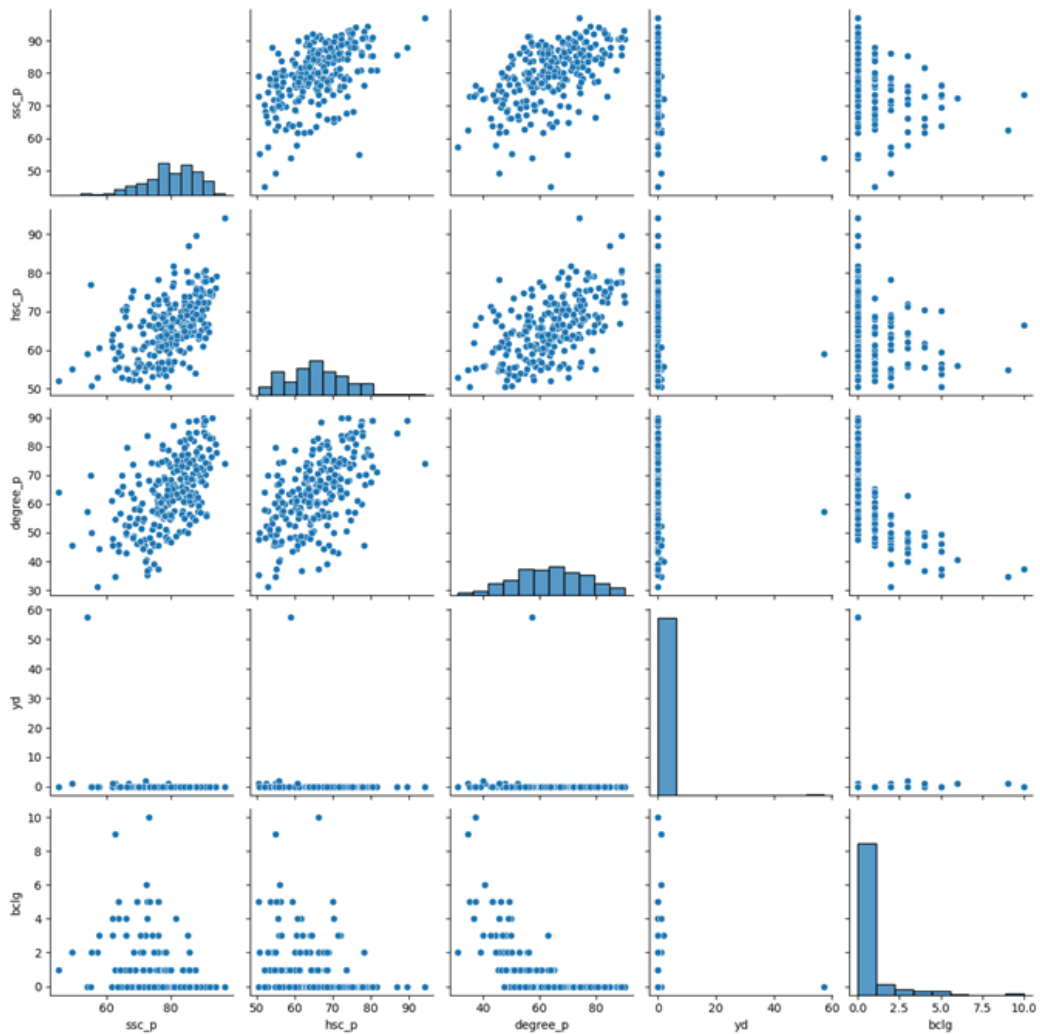


Figure 4 figure showing weights of all features and their importance.

II. RESULTS AND DISCUSSION

The study aimed to predict students' placement success using Logistic Regression machine learning algorithm. The dataset used for the analysis was split into two parts, with two-thirds as the training set and one-third as the testing set.

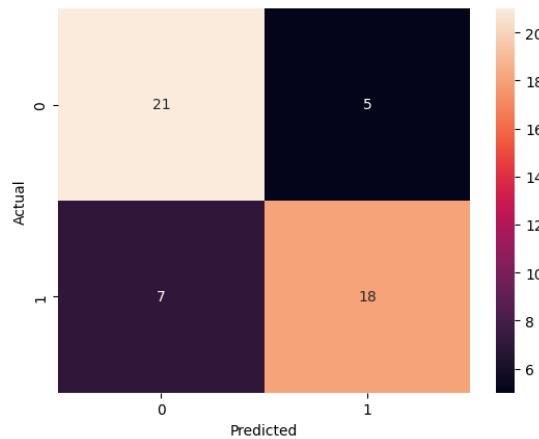


Figure 5 confusion matrix of logistic regression algorithm

The accuracy of the prediction model was calculated using the formula:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} * 100.$$

where TP, TN, FN, FP represents the number of true positives, true negative, false negative and false positive cases.

The results showed that Logistic regression algorithm had accuracy with 76.5 %. The study also analyzed the impact of input variables on the output variable and compared the efficiency of the three approaches in terms of accuracy, recall, and precision.

TABLE II. EFFICIENCY OF LOGISTIC REGRESSION PLACEMENT PREDICTION MODEL

Accuracy	Precision	F1	Recall
76.5 %	78.3 %	75 %	72 %

CONCLUSION

The campus placement activity is important for both institutions and students. Machine learning algorithms such as Logistic Regression is used to predict the placement status of final year BSc (Computer Science) and BCA students. The accuracy of this algorithm is 76.5% depending on the dataset used. The model predicts student placement status using various educational parameters, which can help students identify their capabilities and improve accordingly. These predictions can also aid academic planning to prepare proper strategies and improve placement statistics for future years. Logistic regression is shown to be an important analytical technique for use in graduate placement analysis.

REFERENCES

1. Placement Prediction and Analysis using Machine Learning, Naresh Patel K M, Dept of CSE, BIET, Davangere India, 2022, International Journal of Engineering Research & Technology (IJERT)
2. Placement Prediction using Various Machine Learning Models and their Efficiency Comparison, Irene Treesa Jose , Daibin Raju, Jeebu Abraham Aniyankunju, 2020, International Journal of Innovative Science and Research Technology
3. STUDENT PLACEMENT PREDICTION USING MACHINE LEARNING, Shreyas Harinath , Aksha Prasad , Suma H S3, Suraksha A, Tojo Mathew, 2019, International Research Journal of Engineering and Technology
4. Student Placement Prediction Using Support Vector machine Algorithm, Dr. Rajiv Suresh Kumar, Fathima Dilsha, Shilpa A N, Sumayya A A, 2021, International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

5. STUDENTS PLACEMENT PREDICTION USING MACHINE LEARNING ALGORITHMS
Dr. Kajal Rai, 2022, South Asia Journal of Multidisciplinary Studies
6. Undergraduate Student's Campus Placement Determination Using Logistic Regression Analysis for Predicted Probabilities on Uncertain Dataset, Chandra Sekhar K, K Santhosh Kumar, International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING