Predicting Student's Campus Placement Probability using Binary Logistic Regression

D. Satish Kumar, Zailan Bin Siri, D.S. Rao, S. Anusha

Abstract: ***Students aspiring for technical education generally select educational institutions with good track record in campus placements. Many a times the reputation of such institute is determined by the pay packages offered by recruiters to its students. In this context it is pertinent to investigate and identify those factors that may influence the student campus placement chances in technical education. The State of Andhra Pradesh which has a high concentration of technical education institutes was chosen as the study area. A careful review of literature lead to the identification of six hypothetical determinants of student campus placement in technical education. A random sample 250 MBA student's placement data were gathered from different institutes and six predictor binary logistic regression model was fitted to the data to estimate the odds for the student campus placement. Estimated Results of the study indicate that the chances of campus placement is influenced by four predictors: CGPA, Specialization in PG, Specialization in UG and Gender.***

Key words: ***Campus Placements Technical Education Odds Ratio Binary Logistic Regression Goodness of Fit Confusion Matrix***

1. INTRODUCTION

Most of the companies are banking on external placements. Many MNC's get high caliber candidates from the existing universities in the country. Those who seek good placements find a chance to get into good companies with respect to their candidates. However, students have to put in their best affords to get into covered positions. Those seeking admissions should do a thorough homework about the company and its placements to do the job. The onus lies on the university to provide good placements to students besides providing them required training. There are several aspects to be reckoned upon both academic and non-academic which hold ground while providing placements.

2. PROBLEM STATEMENT

Campus placement chances are important criteria while selecting an educational institution by the student. Several factors contribute to the campus placements for a student both academic and non-academic. Of these, academic achievements, both present and past, softskills, domain knowledge, area of specialization, socio economic attributes are considered here. The present study attempts to build a classification model which can predict the probability that a student who joins in a management program with certain features is going to be placed or not. The model is based on binary logistic regression which is a commonly used classification algorithm in this type of problems.

3. OBJECTIVES OF THE STUDY

- To study the nature of campus placements which is useful for both Students and Institution.
- To Build a model that can be used to predict the probability that a randomly chosen student will be placed or not.
- To Identity the factors that are influencing the placement chances of a student.in technical education.

4. LITERATURE REVIEW

G.Vadivu, K.Sornalakshmi are studied the machine leaning algorithms K-Nearest neighbour methods KNN and Naïve Bayes are used to predict the employability skill based on their regular performance. Algorithms like KNN and Naïve Bayes are useful to classify the objects into one of several groups based on the values of several variables. Manoj K Shukla, Pranay Rambade, Jay Torasakar, Rakesh Prabhu, and Deepali Maste made a study on Students Placement Prediction Model Using Logistic Regression. Placements are always based on the individual performances of the students at the time of interview. Unfortunately, when the academic performance is a bit low student are not selected which is unfair. Students may be good in aptitude along with technical and communication skills. The study of the classification model for placement is based on classification approach which enables the recruiter to find the write kind of evaluation methods to select students for specified job. Using ANN, Vijay N. Kalbande, et al. made a study on predicting the performance of engineering students in campus placement for IT Sector. In this they identify the different employability skills which are affecting the performance of engineering students in campus placement. They develop An Artificial Neural Network (ANN) to predicting probability of placement based on employability skills. Tripti Mishra, Dharminder, Sangeeta Gupta made a study about model prediction on Student's employability with data mining. The major concern for higher education institutions is employability of stunts. Early prediction of employability of the students is always well-timed action. Hitarthi Bhatt, et al. study on Use of ID3 decision tree algorithm for placement prediction. Whether the student will get placement or not is focal point of this paper. Rama Krishna, Bode Prasad, Satyanarayana Murthy, Improved

Revised Manuscript Received on July 01, 2019.

D. Satish Kumar, Department of Mathematics, Koneru Lakshmaiah Education Foundation, Vaddeswaram- 522502, India

Zailan Bin Siri, Institute of Mathematical Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, MALAYSIA

D.S. Rao, Business School, Koneru Lakshmaiah Education Foundation, Vaddeswaram- 522502, India

S. Anusha, Department of Mathematics, Koneru Lakshmaiah Education Foundation, Vaddeswaram- 522502, India

Decision Tree Based Algorithm used in Placement Prediction Analysis in University. Ajay Kumar Pal, Saurabh Pal are enhanced the method of evaluation to predicting student placements. Sudheep et al. made a study on Data mining Framework in Placement Chance Prediction Problems. Data Mining is such a promising technology whose worth becomes evident when it can be applied to a domain where a common man is benefited.

5. HYPOTHESIS FORMULATION

Based on the literature survey the

H1: Gender has an impact on student placement.

H2: PG specialization has an impact on student's placement.

H3: PG cgpa is positively influencing the student's probability of placement.

H4: UG specialization has an impact on student's placement.

H5: UG cgpa is positively influencing the student's probability of placement.

H6: Softskills competency has positive impact on student's placement.

6. RESEARCH METHODOLOGY & RESULTS

A random sample 250 MBA students placement records from 5 leading institutions were obtained. A six predictor binary logistic regression model was fitted to the data. The dependent variable is binary with outcomes: placed or not-placed. The predictors considered are: CGPA in UG and PG, Specialisation in UG and PG, Soft Skill Score and Gender. Standard diagnostic checks were applied to validate the fitted model. The open source software package R was used to analyze the data.

7. BINARY LOGISTIC REGRESSION MODEL ESTIMATION

Dependent Variable	Placement Status		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.00	0.00 – 0.24	**0.016**
gender 2	4.88	1.96 – 12.96	**0.001**
Pgspn 2	1.23	0.49 – 3.09	0.664
Pgspn 3	1.20	0.34 – 4.24	**0.023**
Pgspn 4	1.98	0.35 – 16.01	0.463
Pgspn 5	4.60	0.36 – 112.74	0.252
Ugspn 2	1.81	0.57 – 6.01	0.318
Ugspn 3	3.40	1.31 – 9.33	**0.014**
Ugspn 4	2.24	0.72 – 7.28	0.169
Ugspn 5	2.06	0.08 – 56.56	0.624
d MBA CGPA	2.83	1.70 – 5.02	**<0.001**
d UG CGPA	0.71	0.41 – 1.23	0.229
d soft skills	0.99	0.92 – 1.06	0.824
Observations	250		

Cox & Snell's. R^2 / Nagelkerke's. R^2 0.195 / 0.23

8. MODEL DIAGNOSTICS:

We have to carry out diagnostic tests before the binary logistic regression model can be validatedNull deviance: 214.74 on 247 degrees of freedom
Residual deviance: 149.62 on 229 degrees of freedom
AIC: 207.62
Hosmer-Lemeshow Goodness-of-Fit Test
Summary: model seems to fit well.
Cox & Snell's. R^2 - 0.195
Nagelkerke's. R^2 - 0.262

8.1 Confusion matrix:

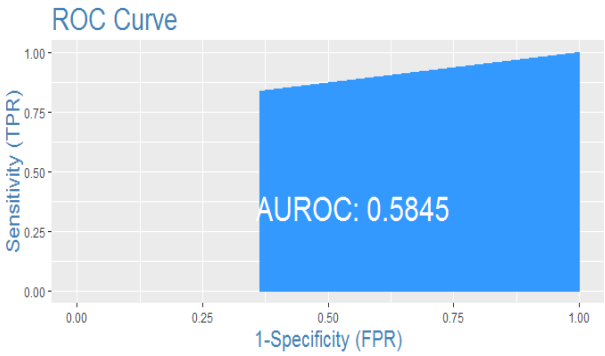
N=257	Predicted: 0	Predicted: 1
Actual: 0	97	45
Actual: 1	24	81

$$\frac{\text{True Positive + True Negatives}}{\text{True Positive + True Negatives + False Positives + False Negatives}}$$

True positive=97
True negative=81
False negative=45
False positive=24

Model accuracy = 178/247
=72.06

The Hosmer-Lemeshow Goodness-of-Fit Test conforms that the model fit for the data. From the Confusion Matrix is the model accuracy is 72% which is a fairly reasonable accuracy level. The Cox & Snell's. R^2 / Nagelkerke's R^2 provides the pseudo R^2 measures for the model. The ROC curve is shown below.



9. DATA ANALYSIS AND INTERPRETATION

Results of the six predictors logistic model were used to test the research hypotheses regarding relationship between the likelihood that a student is placed and the student's Gender, PG specialization, PG CGPA, UG specialization, UG CGPA, Soft skills competency were shown in table. According to the results being displayed in table 1 we can draw the following conclusions.

1. The odds of a student being placed was positively related to his PG CGPA (2.83(p=0.001)), Gender, PG specialization and UG specialization.
2. The higher the PG CGPA the more likely that a student is being placed.

3. Given the same CGPA in PG, male student are more likely to be placed compared to female student.
4. Given the same CGPA student's belonging to Marketing and Finance are more likely to be placed.
5. Given the same CGPA student's from B.Com background in their UG are more likely to be placed.
6. Based on the ROC analysis we may say that the model with the six predictors considered has a 60% accuracy in predicting whether a student gets placed or not.

REFERENCES:

1. G.Vadivu, K.Sornalakshmi, "Applying Machine Learning Algorithms for Student Employability Prediction Using R". *Int. J. Pharm. Sci. Rev. Res.*, Vol.43(1), No. 11, pp. 38-41, 2017.
2. Manoj K Shukla et al. "Study on Students Placement Prediction Model Using Logistic Regression", *International journal of research and technology*, Vol.05(1), pp.1-4,2017.
3. Vijay N. Kalbande et al. "Developing a model to predict employability of engineering students in campus placement for it sector", *International Journal of Advance Research in Engineering, Science & Technology*, Vol. 2(6), pp.201-206, 2015.
4. Tripti Mishra, Dharminder, Sangeeta Gupta, "Study on Student's employability prediction model through data mining", *International Journal of Applied Engineering Research*, Vol 11(4), pp. 2275-2282, 2016.
5. 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*, Vol. 6 (5), pp.4785-4789, 2015.
6. Bode Prasad, T.S. Murthy, "Placement Prediction Analysis in University Using Improved Decision Tree Based Algorithm", *International Journal of Computer Trends and Technology*, vol.17(4), pp.190-195,2014.
7. Ajay Kumar Pal, Saurabh Pal, "Classification Model of Prediction for Placement of Students. I.J.Modern Education and Computer Science", Vol.11, pp.49-56, 2013.
8. Sudheep et al., "A Generalized Data mining Framework for Placement Chance Prediction Problems", "*International Journal of Computer Applications*",Vol.31(3), pp. 40-47,2011.