

# **PREDICTION OF CAMPUS PLACEMENT RESULT USING CLASSIFICATION ALGORITHMS**

## **ABSTRACT**

One of the biggest moments of the students is to get placed in a good company but getting the job isn't easy in the competitive world. There are many factors that decides the student placement results such as academic percentage, skills, university reputation and other factors. Students are facing lots of mental pressure from the placements and their results because they are not aware of which factor makes difference in their interview. So, the aim of the project is to build the prediction model in which we are going to find the important factors in placements and also going to predict whether the student will be placed or not based on the input of the factors. As we are predicting the output of the result which is placed (1) or not placed (0), we are going to use different classification algorithms such as logistic Regression, Decision Tree Classifier, Random Forest and KNN for the dataset and find the algorithm which fits best for this dataset. Accuracy, Precision, Recall, auc-roc curve, F1 score are the evaluation measures that will be taken in order to find the performance of the classification model.

## **INTRODUCTION**

Majority of businesses depend on external placements. Many multinational corporations (MNC's) recruit top talent from the country's established universities. Those who are looking for good placements give you a chance to work for a good company in relation to their candidates. Students, on the other hand, must put in their best efforts to obtain protected positions. Applicants should conduct extensive research prior to applying to do the work about the business and its placements. The burden of proof is on the university to provide students with suitable placements in addition to supplying them with the necessary preparation. There are plenty of aspects to consider, both academic and non-academic, that hold ground while securing placements.

## **RELATED WORKS**

KNN and Naive Bayes algorithms are useful for categorizing objects into one of many classes based on the values of multiple variables. Placements are often made based on the students' individual results during the interview process. Unfortunately, students with poor academic results are not picked, which is unjust. Students can possess strong aptitude as well as technical and communication abilities. The classification model for placement is based on a classification methodology that allows recruiters to find the right kind of assessment methods to choose students for a specific job. Vijay N Kalbande, et al. used ANN to forecast engineering students' success in campus placements for the IT sector. They do so by identifying the various employability skills that influence engineering students' success during campus placement. The employability of students is a significant problem for higher education institutions. Early detection of a student's employability is often a wise step.

## **METHODOLOGY**

We have taken the dataset of campus placement history which has data that is required to build a classification system. We have used four classification algorithms for the dataset to find the best algorithm which suits better for the dataset.

### **1. Dataset**

In our dataset, we have 15 features and 216 rows of data. The Description of features are sl\_no – Serial number, gender – gender=Male/Female, ssc\_p – Secondary education percentage, ssc\_b – Board of Education, hsc\_p – Higher Secondary Education percentage, hsc\_b – Board of education, hsc\_s – Specialization in Higher Secondary Education, degree\_p – Undergraduate Degree percentage, degree\_t – Major in Undergraduate degree, workex – work experience, etest\_p – Employability test percentage, specialization – MBA specialization, mba\_p – MBA percentage, status – placed /not placed, salary – salary offered to candidates.

### **2. Framework**

We have used Jupyter Notebook (Anaconda Navigator) User Interface where we used python to build the classification models. An incredible selection of libraries is one of the primary reasons Python is the most famous programming language utilized for AI. Python Libraries gives base level things so that we don't have to code them from the beginning every time. It lets you access, handle and transform data as ML requires continuous data processing. The libraries or packages we have used for this project are Pandas, Numpy, Seaborn, Sklearn and Plotly.

### 3. Data Preprocessing

The pre-processing done to prepare the data consists of following steps:

1. Loaded the dataset to the pandas data frame
2. Replaced null value rows of salary column.
3. Removing unwanted column as it is not an important feature for our problem.
4. Splitted dependent and independent variable using iloc function.
5. Converting all the string column to numerical column using label encoder
6. Splitting the dataset into training and test data.
7. Min-Max scaling is done before fitting the model so that all the attributes/features will be transformed to the range (0,1).

### 4. Algorithm

We have used four classification algorithm such as Logistic Regression, Decision Tree Classifier, Random tree forest and KNN for this dataset to find the best algorithm.

The parameters used for each algorithm:

Decision Tree Classifier: Criterion, Max\_depth, max\_features, random\_state, splitter.

Random Tree Classifier: n\_estimators, criterion, random\_state.

KNN: n\_neighbors, p.

## EXPERIMENTAL DISCUSSION

In this section, we discuss our dataset split strategy, training, and testing processes that we have implemented in this work and we are also going to cover the strategy of how we work with our dataset, data cleaning, and the different algorithm we used on the dataset to find accuracy.

**Strategy for dividing the dataset:** We randomly divide the dataset into two parts -70% for the training dataset and the remaining 30% for testing purposes. There are almost 215 rows of data in the dataset.

**Data Sanitization and Processing:** We have replaced all the null values in dataset. We have removed the serial number and salary column as it is not the part of the problem and doesn't affect the output of the model. We check the accuracy after removing the columns and there is no change in accuracy. We have splitted the dataset into independent and dependent variable and pre-processed the independent variable with label encoder. Before fitting the data into model, the dataset is splitted and min max scaling is done.

**Training the data:** We have trained the data in all different classification algorithm and got the accuracy of 84.6 % in Logistic Regression, 76.9% in Decision tree classifier, 78.4% in Random Forest and 75.3% in KNN.

	Algorithm Model	Accuracy	Recall	Precision	F1 score	roc_auc
0	Logistic Regression	0.846154	0.978261	0.833333	0.900000	0.752288
1	Decision Tree Classisifer	0.769231	0.891304	0.803922	0.845361	0.682494
2	Random Forest	0.784615	0.913043	0.807692	0.857143	0.693364
3	KNN	0.753846	0.913043	0.777778	0.840000	0.640732

Figure 1. Evaluation Metrics

From the observation, we could say Logistic regression has higher accuracy than other classification algorithm and we have evaluated the Logistic Regression model using other evaluation metrics like recall, precision, f1 score and roc\_auc.



Figure 2. Confusion Matrix of Logistic Regression

From the above figure 2, we could observe the model has 45 true positive, 1 false negative, 10 true negative, 9 false positive. By using the observed values, we could find evaluation metrics.

**Recall** =  $TP / TP + FN = 45 / 45 + 1 = 0.978$

Out of all positive classes, it predicts 97.8% positive classes correctly.

**Precision** =  $TP / TP + FP = 45 / 45 + 9 = 0.833$

Out of all positive classes which we predicted correctly, 83.3 % are actually positive.

**F1 score** =  $2 * \text{precision} * \text{recall} / \text{precision} + \text{recall} = 2 * 0.978 * 0.833 / 0.978 + 0.833 = 0.900$

## ROC Curve

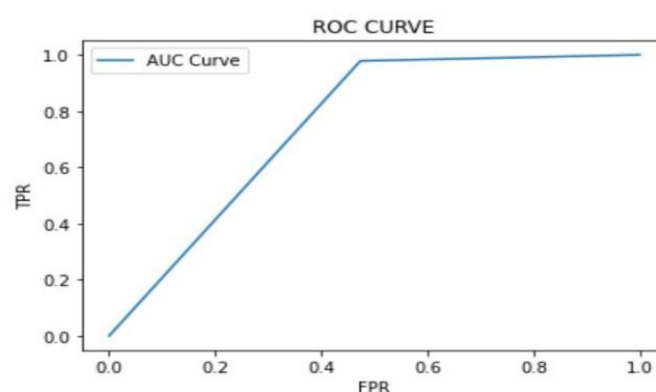


Figure 3. ROC Curve

## CONTRIBUTION

Tasks we distributed amongst the team members are as follows:

- Dataset been identified by **Rashmi** and **Ankita**.
- Finding the algorithms which effects in the prediction done by **Harikarasuriya** and **Nishanth**.

- Data cleaning, null checks, label encoding, evaluating data into chunks and testing scenarios done by **Ankita** and **Rashmi**.
- Perform the activity in Python with the dataset done by **Harikarasuriya, Ankita, Rashmi** and **Nishanth**.
- Applying different Machine learning algorithms like Logistic regression, Decision Tree Classifier, Random forest, k-NN on the datasets been done by **Harikarasuriya** and **Nishanth**.
- Finding the more accurate algorithms amongst all with accuracy, precision, etc done by **Rashmi** and **Ankita**.
- Finalising all these steps together done by the team.

## CONCLUSION

These factors which affects the campus placements we tried using the data in various classification models and found that Logistic Regression is one of the best models we found with the used dataset as its giving good accuracy rate, Recall, Precision and F1 score. We also tried other algorithms as Decision Tree Classifier, Random Forest and KNN, which helped us comparing in between for the accuracy and effectiveness. The advantage with these algorithms that it provides accurate predictions.

## REFERENCES

- Bruce Basta, D' Archy Becker, P.Jane Staly, Richard S Sathe and Kate Mooney(2007), "Effective Campus Recruiting : Faculty perspective", CPA Journal,pp. 62-65
- TimesJobs.com Bureau, "Train campus recruits early to build broader talent base" (Online) Available: [http://articles.economictimes.indiatimes.com/2013-05-06/news/39065498\\_1\\_campus-recruitment-surveyed-organisations-timesjobs-com](http://articles.economictimes.indiatimes.com/2013-05-06/news/39065498_1_campus-recruitment-surveyed-organisations-timesjobs-com) (May 6,2013)
- Callanan, G.andC. Benzing 2004.Assessing the role of internships in the career-oriented employment of graduating college students.Jour. of Education & Training. 46(2): 82-89.
- Beggs, J.M, J.H. Bantham,and S. Taylor. 2008. Distinguishing the factors influencing college students' choice of major. College Student Jour. 42(2): 381.
- <https://www.youtube.com/watch?v=8HriTVx4L7c>
- <https://www.kaggle.com/benroshan/factors-affecting-campus-placement/tasks>
- <https://www.kaggle.com/benroshan/factors-affecting-campus-placement/discussion/177661>