IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Campus Placement Predictive Analysis Using Machine Learning.

Prof.Dr.D.S.Hirolikar
Information Technology
Pune District Education Association
COE Manjari, Bk.
Pune, India

Supriya Gunjal
Information Technology
Pune District Education Association
COE Manjari, Bk.
Pune, India.

Mahesh Bhanwase
Information Technology
Pune District Education Association
COE Manjari ,BK.
Pune ,India.

Mangesh Hagare
Information Technology
Pune District Education Association
COE Manjari ,BK.
Pune, India

Rohit Hedau
Information Technology
Pune District Education Association
COE Manjari ,BK.
Pune, India

Mayur Ulhare
Information Technology
Pune District Education Association
COE Manjari, BK.
Pune, India

Abstract:

A campus placement prediction system is developed to calculate the possibility of a student getting jobs in a company through campus placements. The model takes many parameters that can be used to get an idea about the skill level of the student. While some data are taken from the college level like academic performance, CGPA, pointers, attendance etc. others are obtained from tests conducted in the placement management system. Combining these data points, the model is to accurately predict if the student will or will not be placed in a company. Also, Data from past year students are used for training the model. We are using educational data mining by which we can get real past year's student data of that specific college. This will help the machine learning model to be more effective about the predictions of that particular college.

The purpose behind this placement prediction system is to help students improve their results, and academic performance and also develop another soft skill that will help them to maximize their chance of getting placed. Such a study will help the faculties of the college to train the students accordingly and improve the placement department of their institutions. This will give an idea about how students are doing and will ensure their institution can satisfy the needs of recruiters.

For this supervised machine learning especially logistic regression is better, Logistic model designing plays a key role in getting correct predictions. This process includes the selection of tuples for training data and their pre-known outcome often known as real data.

Keywords:

Data Mining, Campus Placement, Logistic Regression, Machine Learning, Accuracy, Prediction.

Introduction:

NOWADAYS the number of educational institutes is growing day by day. The aim of each higher educational institute is to help their students to get a well-paid job through their placement cell. One of the biggest challenges that higher learning institutes face these days is to uplift the placement performance of scholars. The goal of this system is to predict whether the student will get a campus placement or not based on various parameters such as gender, SSC percentage, HSC percentage, HSC stream, degree percentage, degree type, work experience & e-test percentage. This research focuses on various algorithms of machine learning such as Logistic Regression, Decision Tree, K-Nearest Neighbours and Random Forest in order to produce economical and correct results for campus placement prediction. This system follows a supervised machine learning approach as it uses class-labelled data for training the classification algorithm.

Objective:

The placement Prediction system predicts the probability of an undergraduate student getting placed in a company.

The main objective of this model is to predict whether the student he/she gets placed or not in campus recruitment.

Problem Definition/Statement:

- placement predictor takes many parameters which can be used to assess the skill level of the student.
- While some parameters are taken from the university level, others are obtained from tests conducted in the placement management system itself.
- Combining these data points, the predictor is to accurately predict if the student will or will not be placed in a company. Data from past students are used for the training the predictor.

Literature survey:

In this paper author used UG and PG both student data like MBA CGPA, UG CGPA, soft skills. Their model results from the odds of a student being placed were positively related to his PG CGPA (2.83(p=0.001)), Gender, PG specialization, and UG specialization. The higher the PG CGPA the more likely it is that a student is being placed. 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. Also, Given the same CGPA students belonging to Marketing and Finance are more likely to be placed.[1]

In this paper, author discussed Students Placement Prediction Model Using Logistic Regression. They used previous year placement data as training data which they applied on a logistic mathematical equation. After mapping this data the final logistic model was used to predict the placement chances of next year students with help of thirteen same parameters which were used during model processing. They have also shown the pseudo-code of the whole process.[2]

In this paper author discussed PPS with Logistic regression. Placement and academic data provided by GNDEC are used in this model. Data is cleaned and converted into numbers to use for number crunching. Machine learning technique is used to design and implement a logistic classifier that predicts the probability of the student getting placed. This predictive model predicts the

future outcomes of each student in future sessions of jobs. The parameters are learned by running a gradient descent algorithm on training data that provides an idea about the most necessary features that are responsible for a positive outcome.[3]

In this paper author applied Data mining and knowledge discovery processes to the academic career of students. Which uses an ensemble approach-based voting classifier for choosing best classifier models to achieve better results over other classifiers. Results of the experiment have shown 86.05% accuracy of ensemble-based approach which is better than other classifiers. This prediction has an accuracy of 86.05%, F1-Score 0.86, CohenKappa Score is 0.72, and MSE 0.14 which is better than other classifiers.

Proposed system Architecture Diagram:-

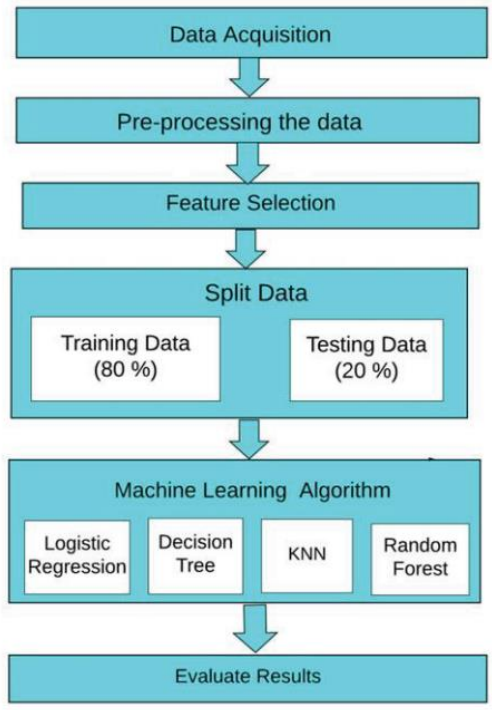

Fig. 1. Architecture Diagram

Methodology

The steps involved in this system are as follows,

A. Data Acquisition:

The campus placement dataset is collected from the Kaggle website. Here is the link for the dataset:https://www.kaggle.com/benroshan/factors-affecting- campus placement? Select=Placement_Data_Full_Class.csv

The dataset consists of various attributes such as Serial Number, Gender, SSC percentage, SSC Board - Central/ Others, HSC percentage, HSC Board, HSC Specialization, Degree Percentage, UG Degree Stream, Work Experience, E –test Percentage, Degree Specialization, Degree Percentage, Placement Status & Salary. The size of dataset is 19.71 KB & it has total 215 records.

1) Handling missing values:

In our dataset, missing values are present only in the salary column as these values correspond to the students who didn't get placed in any placement drive. So it is assumed that the missing values in the Salary Column are Zero & replaced by zero using fillna(0,inplace=True) function in Python.

2) Handling categorical data:

Since we cannot deal with categorical values directly, mapping is done for attributes having categorical values

3) Feature Selection:

Here, various features are visualized to understand their correlation with the target feature.

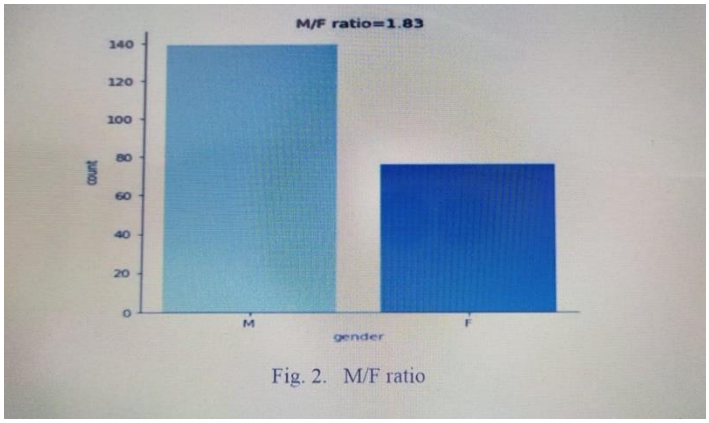

Fig. 2. M/F ratio

4) Split data:

Here, data is divided into two parts i.e. training data & testing data. Where 80% data is taken for training our machine learning algorithm and remaining 20 % data is used for testing whether our trained machine learning model is working correctly or not.

5) Machine Learning Algorithm:

a) Logistic Regression: Logistic regression is a statistical method used to determine the outcome of a dependent variable (y) based on the values of independent variable (x). In our problem dependent variable is placement status and independent variables are the features selected by us in the previous step. This algorithm is mostly used for the problems of binary classification.

```
Confusion Matrix is:
 [[10 10]
 [ 5 40]]


Classification Report is
:              precision    recall  f1-score   support

           0       0.67      0.50      0.57        20
           1       0.80      0.89      0.84        45

    accuracy                           0.77        65
   macro avg       0.73      0.69      0.71        65
weighted avg       0.76      0.77      0.76        65



Accuracy Percent of logisticregression  is:
 76.92307692307693
```

b) Decision Tree: A decision tree is a graph like a tree where nodes represent the position where we select the feature and ask a question, edges represent the answers of the question; and the leaves represent the final output or label of the class.

```
Confusion Matrix is:
 [[20  0]
 [ 2 43]]


Classification Report is
:              precision    recall  f1-score   support

           0       0.91      1.00      0.95        20
           1       1.00      0.96      0.98        45

    accuracy                           0.97        65
   macro avg       0.95      0.98      0.96        65
weighted avg       0.97      0.97      0.97        65



Accuracy Precent decision tree is:
 96.92307692307692
```

c) KNN: K-NN stores all the training data into different classes based on the class labels and classifies new data by checking its similarity with data in the available classes.

```
Confusion Matrix is:
 [[20  0]
 [ 2 43]]


Classification Report is
:              precision    recall  f1-score   support

           0       0.91      1.00      0.95        20
           1       1.00      0.96      0.98        45

    accuracy                           0.97        65
   macro avg       0.95      0.98      0.96        65
weighted avg       0.97      0.97      0.97        65



Accuracy Precent  KNN is:
 96.92307692307692
```

d) Random Forest: Random Forest classifier consists of a number of decision trees which apply on different subsets of our dataset and the average of outputs of all the decision trees is taken to improve the accuracy of output prediction.

```
Confusion Matrix is:
 [[18  2]
 [ 1 44]]


Classification Report is
:              precision    recall  f1-score   support

           0       0.95      0.90      0.92        20
           1       0.96      0.98      0.97        45

    accuracy                           0.95        65
   macro avg       0.95      0.94      0.95        65
weighted avg       0.95      0.95      0.95        65



Accuracy Precent of Rnadom Forest Classifier is:
 95.38461538461539
```

6) Evaluate results:

Accuracy is calculated by following the formula,

Accuracy = (TP + TN) / (TP + FP + TN + FN)

Where

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

Model	Accuracy
Logistic Regression	76.92%
Decision Tree	96.92%
KNN	96.92%
Random Forest	95.38%

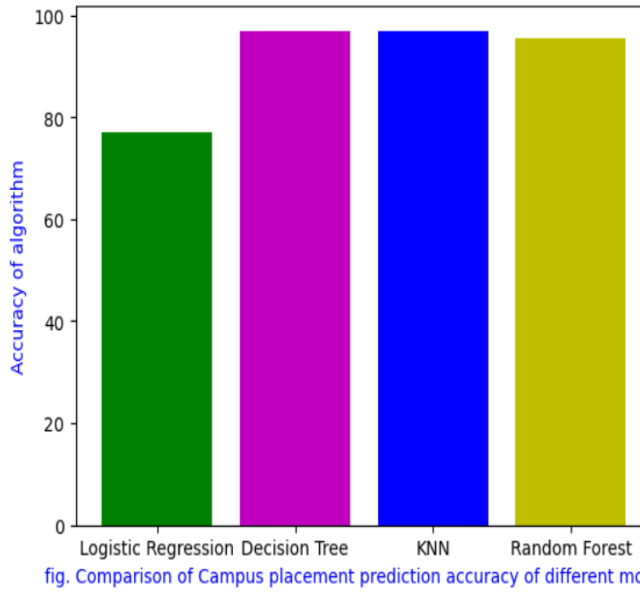

fig. Comparison of Campus placement prediction accuracy of different mo

Conclusion

TPO's manual student placement class prediction is a challenging task. Utilizing data mining, we can fix this issue and aid in the placement of students' predictions. Our recommended system uses a technique for predicting student placement that predicts a certain student's placement using algorithms. predicting the campus placement of students based on his/her performance in academics and other skills and exams. The model is trained using the previous year's data like students marks in all semesters, personal interests, internships, other activities, etc.

FUTURE SCOPE

Accuracy may further increase by application of more advanced techniques such as deep learning & experimenting with different activation functions of neural networks such as linear, sigmoid, tanh & ReLU. We can also experiment with different cross validation techniques such as 3 Fold, 5 Fold, 10 Fold, 15 Fold cross validation in order to analyze the change in accuracy.

References:

1] Irene Treesa Jose, Daibin Raju, Jeebu Abraham Aniyankunju, Joel James, Mereen Thomas Vadakkel "Placement Prediction using Various Machine Learning Models and their Efficiency Comparison" Volume 5,P 1007,1008 Issue 5, May – 2020

[2] D. Satish Kumar, Zailan Bin Siri, D.S. Rao, S. Anusha "Predicting Student's Campus Placement Probability using Binary Logistic Regression" Volume-8 Issue-9, July, 2019

[3] Manoj K Shukla, Pranay Rambade, Jay Torasakar, Rakesh Prabhu, Prof. Deepali Maste "Students Placement Prediction Model Using Logistic Regression" Volume 5, Issue 01

[3] Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar "PPS - Placement Prediction System using Logistic Regression" P 338,339 2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE)

[4] Ravi Kumar Rathore and J. Jayanthi, "Student Prediction System For Placement Training Using Fuzzy Inference System", ICTACT Journal On Soft Computing,2017

[5] 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.

[6] Bharambe, Yogesh, "Assessing employability of students using data mining techniques", Advances in Computing, Communications and Informatics (ICACCI), 2017.

[7] P. Khongchai, P. Songmuang, "Random Forest for Salary Prediction System to Improve Students Motivation", Signal- Image Technology & Internet-Based Systems (SITIS), 2016.

[8] 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.

[8] 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.

(IJERT), 2016.