JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Placement Prediction Model Using Ensemble Learning

Jyoti Bhanwase[1], S. S. Palnitkar[2] ,Gunjan Gahukar[3], Tushar Budhe[4]

E&TC Department, SKNCOE, SPPU Pune, India

bhanwasejyoti@gmail.com[1]
smita.palnitkar_skncoe@sinhgad.edu[2]
gunjangahukar967@gmail.com[3]
tusharbudhe0110@gmail.com[4]

***Abstract*- Every academic enthusiast's main goal is to get placed in a reputable MNC, and even an institute's reputation and annual admittance depends on the placement it delivers to its students. As a result, any system that predicts student placements will have a favourable influence on an institute, increasing strength and reducing workload in the training and placement office (TPO). Machine Learning techniques can be used to extract knowledge from previously placed students and predict the placement of future students. The data for training comes from the same institute that does the placement prediction. Along with the feature selections, appropriate data pre-processing procedures are used. Some domain knowledge is utilized. We employed Logistic, SVM, KNN, Decision Tree, Random Forest, advanced techniques like Bagging and Boosting to reach 74 percent in AdaBoost Classifier.**

***Keywords* -** Pre-processing, Feature Selection, Domain expertise, Outliers, Bagging, Boosting, SVM, KNN, Logistics

I.INTRODUCTION

Nowadays, placement is very significant in the world. Unemployment is rampant. The ranking and evaluation of Institutes are based on the average package amount and They provide a significant amount of placement. The main goal of this model is to forecast whether or not The student may or may not be placed. Various types of Logistic regression, SVM, and other classifiers were used. Gradient, Decision Tree, Random Forest, KNN, AdaBoost XG-Boost and boosting For this, academicians from all over the world The students are taken into account. As the placements go, occur throughout the final academic year, hence final semesters are not taken into account

II. LITERATURE SURVEY

Various studies and students have published relevant work in national and international research papers, this is to comprehend the purpose, types of algorithm they have used, and various pre-processing techniques, Feature. Random Forest and Decision Tree were employed by Pothuganti Manvitha and Neelam Swaroopa (2019). After examination, the accuracy of the Decision Tree is 84 percent and the Random Forest is 86 percent. As a result of the aforesaid analysis and prediction, it is preferable to utilise the Random Forest method to predict placement results [1]. Senthil Kumar Thangavel, Divya Bharathi P, and Abijith Sankar (2017) employed Decision Tree, Logistic Regression, Metabagging Classifier, and Nave Bayes to get the maximum accuracy of 84.42 percent in Decision Tree. The goal of the study is to forecast the placement status of Btec[2]. Ajay Kumar Pal and Saurabh Pal (2013) are using three Weka-based classification algorithms to predict student placement after completing MCA. The most accurate algorithm based on the placement data is Nave Bayes Classification, which has an accuracy of 86.15 percent and takes 0 seconds to create. In comparison to the others, the Nave Bayes classifier has the lowest average error of 0.28. [3]. Mangesh Kolhal, Syed Ahmed, Aditya Zade, Shubham Gore, Prashant Gaikwad (2017). Their goal is to look at previous year's student data and anticipate current students' placement chances as well as the institution's percentage placement possibility. They employed the C4.5 Decision Tree Algorithm. The model is generated by applying decision tree C4.5 algorithms to the company's previous year data and current requirements, and this model may be used to predict students' eligibility in various companies. According to the company's eligibility standards, they will notify those candidates who are eligible for the campus interview and will assess candidate eligibility based on percentage and technology [4]. Deeksha K C, Vishal Prajwal R, Vrushak K, Nandini M S, Apoorva Rao r (2018). They used clustering techniques as well as the Nave Bayes algorithm to divide students into five separate categories: dream company, core company, mass recruiters, not eligible, and not interested. [5]

III. PROBLEM STATEMENT

The placement prediction model simply examines students' academic achievement in order to forecast whether or not they will be placed. We cannot judge students' placement just on the basis of their academic achievements, because some students may excel in aptitude, technical, and communication abilities despite their academic shortcomings. Parameters such as cgpa, logical, and technical skills are needed to determine a student's placement. Academic achievement is crucial, but the algorithm is built to forecast placements based on the students' parameters.

IV. PROPOSED SYSTEM

Fig1. Flow chart

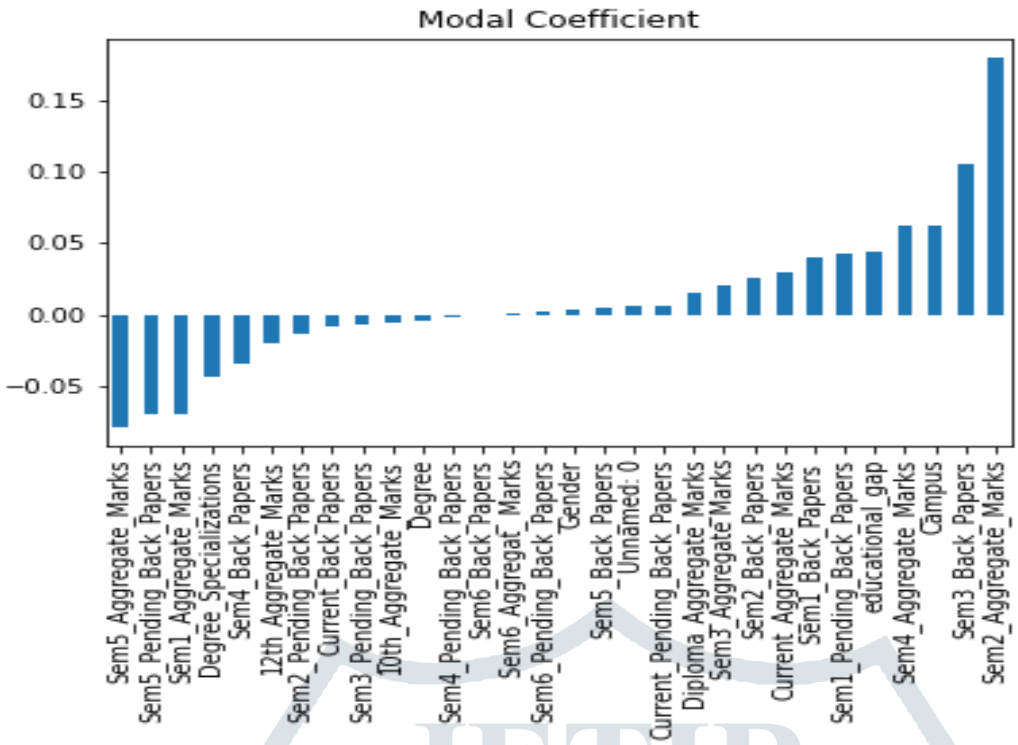
4.1 DATA GATHERING AND PRE-PROCESSING

The data was gathered from Kaggle which included all Bachelor of Engineering (BE) students from three distinct colleges on campus. There are 2338 records in all, with 25 distinct attributes.

i. Students' academic information is contained in this dataset. We amalgamated 12th and diploma marks and established a single column for both because some students have completed their 12th and some are from diploma backgrounds who have directly obtained admission to the second year.
ii. We removed some of the tuples with M.tech backgrounds, and we also removed the NA values from the "current aggregate" column because the entire row was NA.
iii. All NA values in columns "Current Back Papers," "Current Pending Back Papers," and "Sem Back Papers," "Sem Pending Back Papers" were replaced with 0 because it was null only if that student had no backlogs.
iv. The labels of columns "'Degree Specializations", "Campus", "Gender", "year down", and educational gap" were encoded using Label Encoder from the Pre-processing API in SK-learn.

4.2 Feature Selection

We have obtained several outcomes using machine learning Feature Selection methods such as "Ridge," "Lasso," "RFE," "plot importance," "F1 score," and "feature importance."



V. RESULT AND DISCUSSION

Algorithms	Accuracy
Logistic Regression	63.58%
Support Vector Machine	63.28%
KNN	65.49%
Decision Tree	68.86%
Random Forest	74.22%
AdaBoost(DT)	74.00%
Gradient Boosting	75.54%
Extra Tree Classifier	73.34%
DT+ AdaBoost+ Bagging	94.49%

Fig2 . Algorithm performance comparison

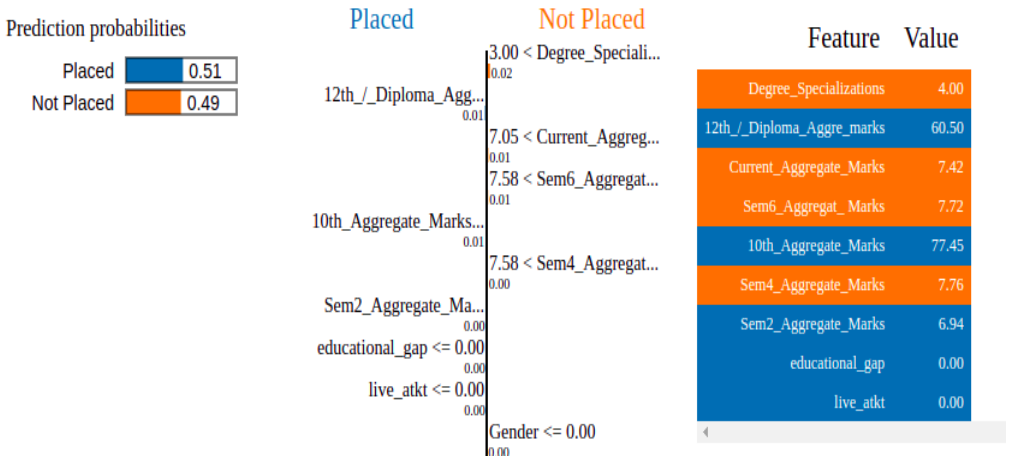


Fig 3. Local Interpretable Model-agnostic Explanations (LIME)

CONCLUSION: We evaluated numerous academic records as well as the aggregate of all semesters, live backlog, dead backlog, education gap, and year down in this model. This model will assist teachers in determining whether or not a kid will be placed prior to the third year so that they may pay special attention to those pupils who are anticipated to not be placed. Even the institute can take significant steps to improve those pupils' qualities prior to their final placement. Various methods were used, but the final model was based on the AdaBoost classifier with Bagging and Decision Tree as the Base Classifier because of its high accuracy .We can add new colleges' datasets to it for prediction if the existing dataset is just for three institutions. We will use Deep Learning algorithms in the future, which may provide superior accuracy than Machine Learning models.

ACKNOWLEDGEMENT

It gives us great pleasure in presenting the preliminary project paper on 'Placement Prediction Model Using Ensemble Learning'. We would like to take this opportunity to thank my internal guide Prof. S.M. Ingawale and co-guides Prof. S.S Palnitkar , Prof. N.S Nikam for giving us all the help and guidance we needed. We are really grateful to them for their kind support. Their valuable suggestions were very helpful. We are also grateful to Dr. S.K. Jagtap, Head of E&TC Engineering Department, SKNCOE for her indispensable support, suggestions. We would also thank our Principal Dr. A.V. Deshpande for his great insight and motivation. Last but not least, we would like to thank my fellow colleagues for their valuable suggestions.

References

[1] Pothuganti Manvitha, Neelam Swaroopa "Campus Placement Prediction Using Supervised Machine Learning Techniques" International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Sept 2019

[2] Senthil Kumar Thangavel, Divya Bharathi P, Abijith Sankar "Student Placement Analyzer: A Recommendation System Using x Machine Learning" 2017 International Conference on Advanced Computing and Communication Systems (ICACCS -2017), x Coimbatore, INDIA, Jan. 06 – 07, 2017

[3] Ajay Kumar Pal, Saurabh Pal "Classification Model of Prediction for Placement of Students" I.J.Modern Education and Computer Science, 2013, 11, 49-56 Published Online, 11 November 2013

[4] Syed A0068med, Aditya Zade, Shubham Gore, Prashant Gaikwad, Mangesh Kolhal "Smart System for Placement Prediction using Data Mining" International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653, Dec 2017

[5] Apoorva Rao r , Deeksha K C , Vishal Prajwal R , Vrushak K, Nandini M S "Student placement analyzer: a recommendation system using machine learning" ijariieissn(o)-2395-4396, Jan 2018

[6] Ravi Tiwari and Awadhesh Kumar Sharma, "A Data Mining Model to Improve Placement", International Journal of Computer Applications (0975 – 8887) Volume 120 – No.12, June 2015

[7] Mangasuli Sheetal B, Prof. Savita Bakare "Prediction of Campus Placement Using Data Mining Algorithm Fuzzy logic and K nearest neighbour" International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 6, June 2016.

[8] Student Placement Prediction Using ID3 Algorithm Namita Puri , Deepali Khot , Pratiksha Shinde, Kishori Bhoite , Prof. Deepali Maste , IJRASET, Volume 3 Issue III, 2015