International Journal of Computer Applications Technology and Research
Volume 8–Issue 09, 358-362, 2019, ISSN:-2319–8656

Campus Placement Analyzer: Using Supervised Machine Learning Algorithms

Shubham Khandale
Student, M.Sc. (Big Data Analytics)
School of Computer Science, Faculty of Science
MIT-WPU, Pune, Maharashtra, India

Sachin Bhoite
Assistant Professor
School of Computer Science, Faculty of Science
MIT-WPU, Pune, Maharashtra, India

Abstract -- The main aim of every academia enthusiast is placement in a reputed MNC's and even the reputation and every year admission of Institute depends upon placement that it provides to their students. So, any system that will predict the placements of the students will be a positive impact on an institute and increase strength and decreases some workload of any institute's training and placement office (TPO). With the help of Machine Learning techniques, the knowledge can be extracted from past placed students and placement of upcoming students can be predicted. Data used for training is taken from the same institute for which the placement prediction is done. Suitable data pre-processing methods are applied along with the features selections. Some Domain expertise is used for pre-processing as well as for outliers that grab in the dataset. We have used various Machine Learning Algorithms like Logistic, SVM, KNN, Decision Tree, Random Forest and advance techniques like Bagging, Boosting and Voting Classifier and achieved 78% in XGBoost and 78% in AdaBoost Classifier.

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

1. INTRODUCTION

Nowadays Placement plays an important role in this world full of unemployment. Even the ranking and rating of institutes depend upon the amount of average package and amount of placement they are providing.

So basically main objective of this model is to predict whether the student might get placement or not. Different kinds of classifiers were applied i.e. Logistic Regression, SVM, Decision Tree, Random Forest, KNN, AdaBoost, Gradient Boosting and XGBoost. For this all over academics of students are taken under consideration. As placements activity take place in last year of academics so last year semesters are not taken under consideration

2. RELATED WORK

Various researches and students have published related work in national and international research papers, thesis to understand the objective, types of algorithm they have used and various techniques for pre-processing, Feature.

Pothuganti Manvitha, Neelam Swaroopa (2019) used Random Forest and Decision Tree. The accuracy obtained after analysis for Decision tree is 84% and for the Random Forest is 86%. Hence, from the above-said analysis and prediction, it's better if the Random Forest algorithm is used to predict the placement results [1].

Senthil Kumar Thangavel, Divya Bharathi P, Abijith Sankar(2017) used Decision Tree, Logistic Regression, Metabagging Classifier, Naïve Bayes and obtain highest 84.42% accuracy in Decision Tree. The objectives, which is to predict the placement status the students in Btech are most likely to have at the end of their final year placements. The accuracy of 71.66% with tested real-life data indicates that the system is reliable for carrying out its major objectives, which is to help teachers and placement cell[2].

Ajay Kumar Pal, Saurabh Pal (2013) they are predicting the placement of student after doing MCA by the three selected classification algorithms based on Weka. The best algorithm based on the placement data is Naïve Bayes Classification with an accuracy of 86.15% and the total time taken to build the model is at 0 seconds. Naïve Bayes classifier has the lowest average error at 0.28 compared to others.[3]

Syed A0068med, Aditya Zade, Shubham Gore, Prashant Gaikwad, Mangesh Kolhal (2017). Their objective is to analyze the previous year's student's historical data and predict placement chance of the current students and the percentage placement chance of the institution. They have used the Decision tree C4.5 Algorithm. Decision tree C4.5 algorithms are applied to the Company's previous year data & current requirement to generate the model and this model can be used to predict the students' eligibility in various companies. According to company eligibility criteria, they will send the notification to those candidates who are eligible for that campus interview and check the eligibility of candidate on the basis of percentage & technology [4].

Apoorva Rao r , Deeksha K C , Vishal Prajwal R , Vrushak K, Nandini M S (2018). They have used techniques like clustering along with that they have used classification rule Naïve Bayes algorithm that will classify students in five

different status i.e. Dream company, Core Company, Mass recruiters, Not eligible and Not interested[5]

3. DATASET DESCRIPTION AND SYSTEM FLOW

This approach was followed in following Figure 3.

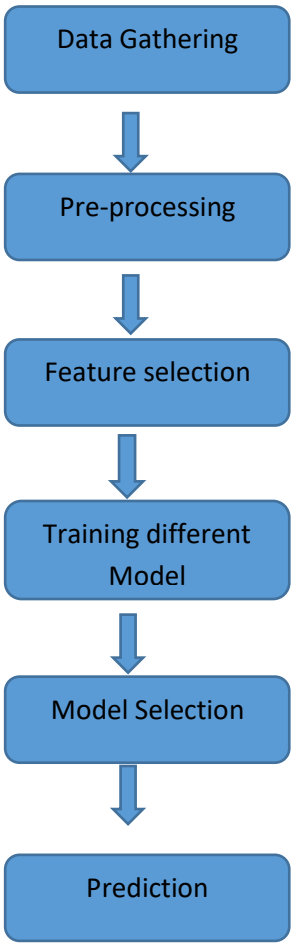

Figure 3. Flow chart

3.1 Data gathering and Pre-processing

The Data was collected from Training and placement department of MIT which consist of all the students of Bachler of Engineering (B.E) from 3 different colleges of their campus. The Data consists of 2338 records with 31 different attribute.

- Dataset contains academic information of students. As some students have completed their 12th and some of them are from diploma background who have directly taken admission to the second year so, we have merged 12th and diploma marks and made a single column for both.
- Some of the tuples where from M.tech background so we have dropped them and even in "current_aggregate" column we have dropped the NA values because the whole row was having NA.
- Replaced all NA values in columns "Current_Back_Papers", "Current_Pending_Back_Papers", all semester wise "Sem_Back_Papers", "Sem_Pending_Back_Papers" with 0 because it was null only if that student have no backlogs
- Using LabelEncoder from Preprocessing API in sklearn encoded the labels of columns "'Degree_Specializations", "Campus"," Gender", "year_down", "educational_gap"

3.2 Feature Selection

As per machine learning Feature Selection algorithms like "Ridge", "Lasso", "RFE", "plot importance", "F1 score" and "feature importance" we have got various outputs

- "Feature importance" with DT

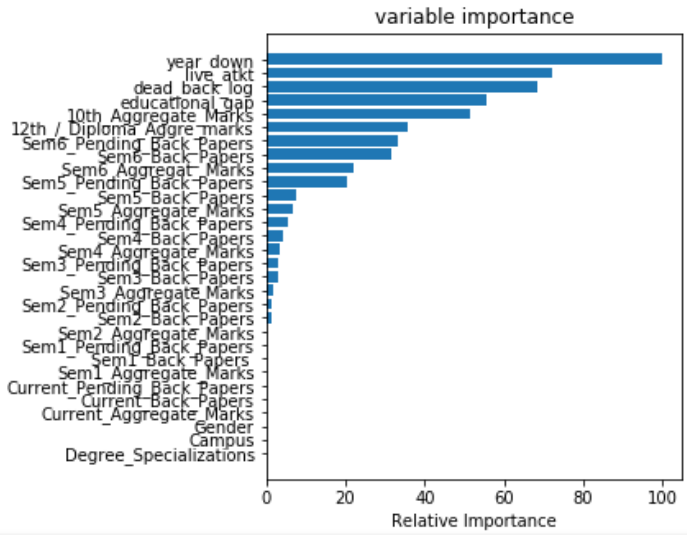

Figure 3.2.1 Feature importance with DT

- "Feature importance" with Random Forest

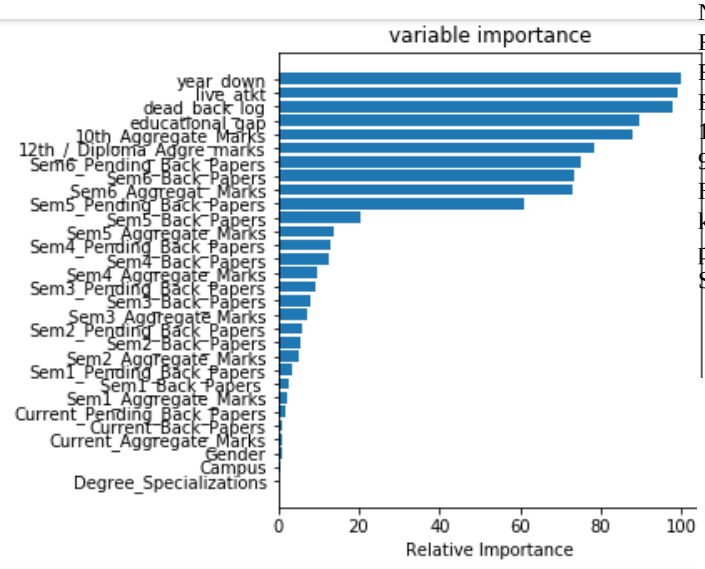

Figure 3.2.2 Feature importance with Random Forest

- "F1 score"

Feature Names	F1 score
Sem4_Aggregate_Marks	312.063809
Current_Aggregate_Marks	286.086537
Sem3_Aggregate Marks	255.771833
Sem2_Aggregate_Marks	164.183078
12th_/_Diploma_Aggre_marks	142.208129
Sem1_Aggregate_Marks	139.183936
Sem6_Aggregat_ Marks	136.333959
Sem5_Aggregate_Marks	131.988165
10th_Aggregate_Marks	128.526784
Sem6_Back_Papers	128.526784
live_atkt	47.908927
Sem5_Back_Papers	45.382049
Sem4_Back_Papers	43.547352

- **"RFE"**

Num Features: 5 Features support : [False False False False False False False False True False False True False False False True False True False False False False True False False False False False False] Features Ranking [25 6 4 3 24 8 13 22 1 10 11 1 23 17 2 1 19 1 5 18 7 21 1 26 16 20 15 12 14 9] selected Features:['Sem1_Pending_Back_Papers','Sem2_Pending_Back_Papers','Sem4_Aggregate_Marks','Sem4_Pending_Back_Papers', 'Sem6_Back_Papers']
Selected features index: [8, 11, 15, 17, 22]

- **"Ridge"**

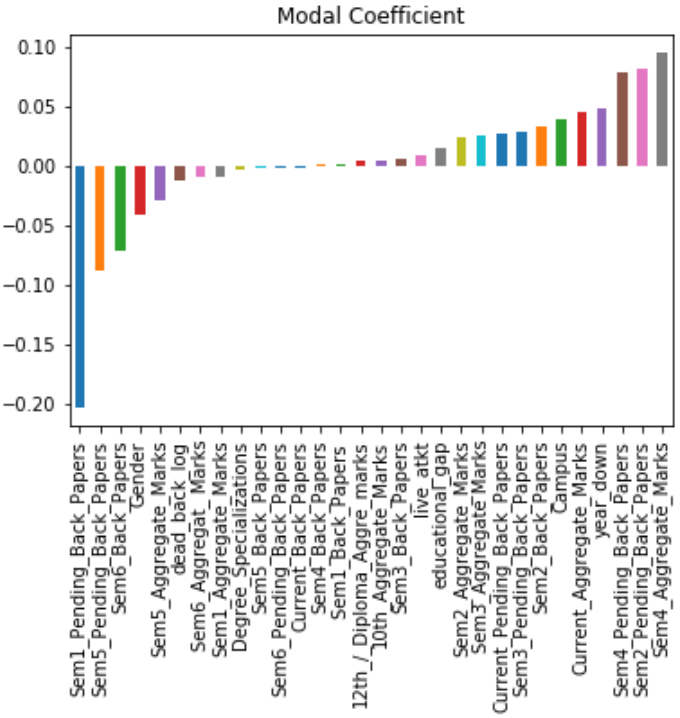

Figure 3.2.3 Feature Selection Using Ridge

- **"Lasso"**

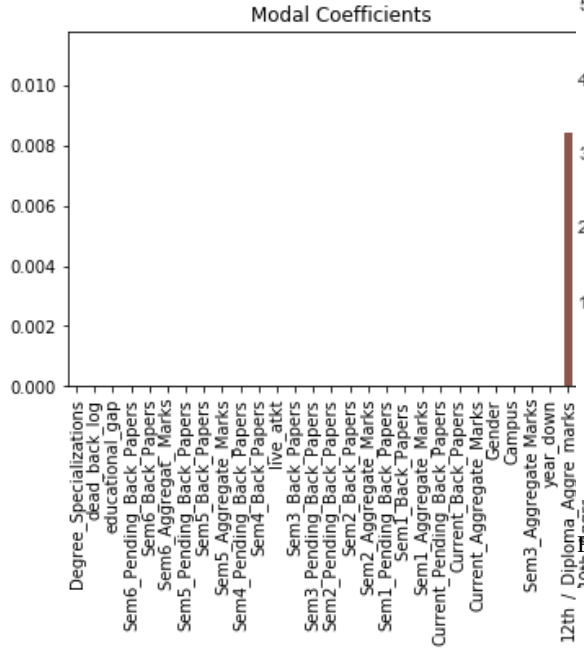

Figure 3.2.4 Feature Selection Using Lasso

But as per the domain knowledge we have selected all the features which are importance for our model

4. EXPLORATORY DATA ANALYSIS

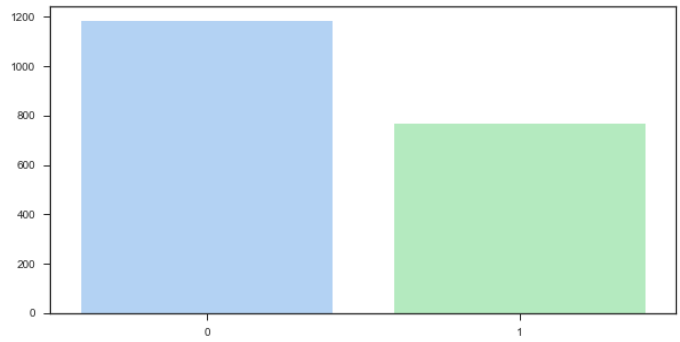

Figure 4.1 Total number of student placed

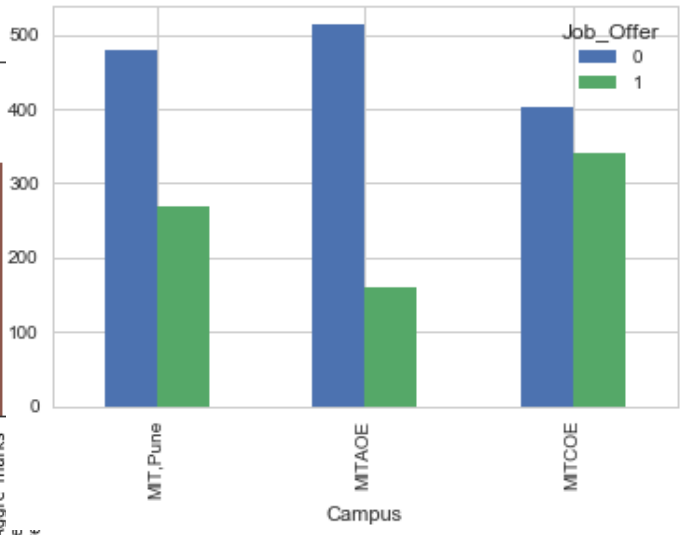

Figure 4.2 Campus wise number of students who got placed

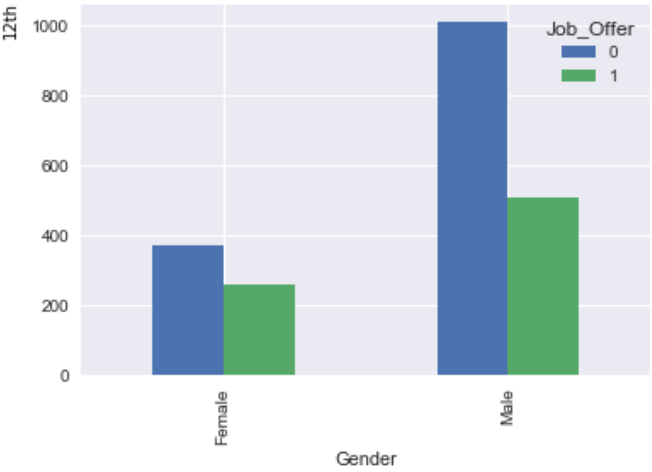

Figure 4.3 Gender Wise Student Placement

5. BAGGING AND BOOSTING

Bagging is nothing but bootstrap aggregating, it is an ensemble method to improve the accuracy and stability of the models. Random samples are taken with replacement and with every new sample that is generated is trained and the ensemble can make a prediction for the new instance by simply aggregating the prediction of all predictors

Boosting is nothing but the ensemble method that can combine different weak learner into a strong learner. Its main aim is to train predictors sequentially. Most popular are AdaBoost and Gradient Boosting.

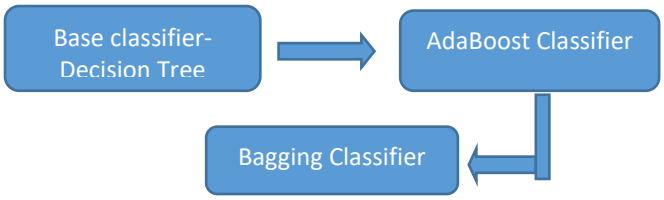

Figure 5.1 Layering of Classifiers

We have used Base Classifier as Decision Tree, over that we have used AdaBoost Classifier and over that we have used Baagging Classifier because we want to tune the accuracy of the model

6. RESULT AND CONCLUSION

Algorithms	**Accuracy**
Logistic Regression	58%
Support Vector Machine	69%
KNN	63.22 %
Decision Tree	69%
Random Forest	75.25%
AdaBoost(DT)	77%
Gradient Boosting	77%
Voting Classifier Soft	69.11%
Voting Classifier Hard	68.43%
XGBoost	78%

In this model, we have considered various academics records along with all semester's aggregate, live backlog, dead backlog, education gap, year down. This model will help the teachers to find whether the student will get placement or not prior in 3rd year only so that they can pay special attention to those students who are predicted as not getting placement. Even the institute can take major steps to improve the qualities of those students before their final placement. Various algorithms were used but the final model is selected on AdaBoost classifier along with the Bagging and Decision Tree as Base Classifier as its accuracy is very high.

The existing dataset was only for 3 colleges further even we can add more college's dataset to it for prediction. In future, we are going to implement Deep learning algorithms which may give better accuracy then Machine Learning models

7. 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 Machine Learning" 2017 International Conference on Advanced Computing and Communication Systems (ICACCS -2017), 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" ijariie-issn(o)-2395-4396, Jan 2018