Students Placement Prediction Using Machine Learning Algorithms

Conference Paper · March 2022

CITATIONS
3

READS
6,654

1 author:

Kajal Rai Saraswat
G L Bajaj
14 PUBLICATIONS **218** CITATIONS

SEE PROFILE

ISSN:2395-1079

Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619

South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

STUDENTS PLACEMENT PREDICTION USING MACHINE LEARNING ALGORITHMS

Dr. Kajal Rai
Assistant Professor Chandigarh University

ABSTRACT

Placements include limitless importance for college students and academic organizations. It helps a student to assemble a robust basis for the expert profession beforehand in addition to a virtuous placement file affords an aggressive side to a school or college inside the schooling arcade. Machine learning is a way of statistical evaluation that automates analytical version construction. This paper makes a specialty of a machine that forecasts if a pupil might be located or now no longer primarily based totally on the pupil's qualifications, ancient statistics, and experience. This forecaster makes use of three machine learning algorithms, namely, Decision Tree, Naïve Bayes, and Random Forest to expect pupil's placement after which evaluation of those algorithms are performed on the idea of accuracy achieved.

KEYWORDS: Data Analysis, Decision Tree, Machine Learning, Naïve Bayes, Placements, Prediction Models, Random Forest.

INTRODUCTION

In the modern-day age, campus placement clutches prodigious importance for college students and academic organizations. While it helps a learner in edifice a strong basis for the expert professionals in advance without going through the real-international activity fight, peer-opposition or own circle of relatives pressure, a first-rate placement report offers an aggressive side to an institute or college with inside the studying marketplace.

Campus placements offer the scholars a foot-in-the-door opportunity, permitting them to start their profession proper when they have finished their direction curriculum. Furthermore, they get to engage and interact with the enterprise specialists in the course of the location drives, which in addition assist to lay a basis for his or her potential profession with inside the destiny as they familiarize with capability contacts from their selected professional field.

Placements have progressively come to be a vital part of an institute's offerings, which turned into now no longer the situation earlier. Nowadays, college students pay unique interest to placement information whilst choosing a university or college for admission.

Machine learning is a growing technology that enables computers to learn automatically from past data.

ISSN:2395-1079
Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619
South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommendation system, and many more.

The dataset considered for this work is the MBA students' data at Jain University, Bangalore of the year 2020 [1]. It includes various attributes like secondary and higher education percentage and board; it also consists of specialization and work experience and many more. With the help of machine learning algorithms, the prediction of students who got placed in the company at campus placements based on various attributes was done in this paper.

LITERATURE REVIEW

In [2] the authors used decision tree and random forest to classify the dataset of campus placed and non- placed students. The accuracy obtained by the decision tree was 84% and the random forest was 86% in this paper.

This paper presents a recommendation framework that forecasts the scholars to have one of the five placement statuses, namely, Dream Company, Core Company, Mass Recruiters, Not Eligible, and Not Interested in Placements. This model benefits the placement cell within an institute to recognize the potential students and pay consideration to and advance their technical as well as social abilities.

In this paper, the authors proposed some supervised machine learning classifiers which can be utilized to foresee the placement of a student in the IT industry centered on their academic performance in class tenth, twelfth, graduation, and backlog till date in graduation. Numerous factors utilized to associate and examine the outcomes of distinctive established classifiers are accuracy score, percentage accuracy score, confusion matrix, heat map, and classification report. Classification report created through advanced classifiers consists of parameters precision, recall, f1-score, and support. The classification algorithms Support Vector Machine, Gaussian Naive Bayes, K-Nearest Neighbor, Random Forest, Decision Tree, Stochastic Gradient Descent, Logistic Regression, and Neural Network are cast-off to develop the classifiers.

The authors in paper [5] used Knowledge Discovery and Data Mining (KDD) which is the placement class procedure using the classification technique. In the primary experimentation, the accurately classified cases were 84.2%, and in the second experimentation using the same data and attributes, provide the best percentage of accuracy as 92.1%. The best outcomes were obtained using Naive Bayes and SMO.

METHODOLOGY

In this paper, the research methodology used can be depicted by the following figure.

ISSN:2395-1079
Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619
South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

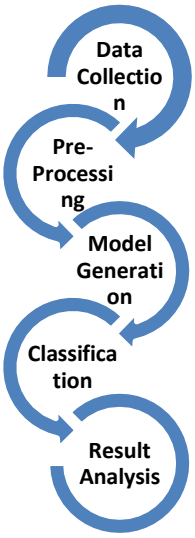

Fig. 1: Methodology Used for Research

DATA COLLECTION

The sample data has been collected from Bangalore college of MBA students. The dataset consists of 215 instances of students.

PRE-PROCESSING

Data pre-processing is a method that is used to convert raw data into clean data which can be used for model construction. This research work consists of pre-processing tasks such as attribute selection, cleaning missing values, and splitting the dataset into training and testing. Some attributes such as serial no. and salary are removed as they do not contribute to classification.

MODEL GENERATION

Three machine learning classifier models are generated using the given dataset. These three models are decision tree, naïve bayes, and random forest. These models are used to classify the dataset based on various selected attributes.

CLASSIFICATION

Classification is done to predict the students who are placed in some company on campus and those who are not placed. To classify the students three well-known machine learning classifiers are used in this paper.

RESULT ANALYSIS

The results obtained from different classifiers are analyzed and compared based on accuracy.

EXPERIMENTATION

Weka is open-source software that implements a large collection of machine learning algorithms and is widely used in data mining applications. For experiments dataset downloaded from Kaggle was in csv file. This file was loaded into WEKA explorer. The classify panel enables the user to apply classification and regression algorithms to the resulting dataset, to estimate the accuracy of the resulting predictive model, and to visualize erroneous predictions, or the model itself. The algorithm used for classification is Naive Bayes, Decision Tree, and Random Forest. Under the "Test options", the 10-fold cross-validation is selected as our evaluation approach. Since there is no separate evaluation data set, this is necessary to get a reasonable idea of the accuracy of the generated model. This predictive model provides a way to predict whether

ISSN:2395-1079
Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619
South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

a new student will place or not in an organization.

UPLOAD FILE IN WEKA

The following figure shows the sample from the dataset which is uploaded in weka for experimentation. There are various attributes in this file such as gender, work experience, specialization, senior secondary percentage, board of senior secondary, etc.

gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisat	mba_p	status
M	67	Others	91	Others	Commerc	58	Sci&Tech	No	55	Mkt&HR	58.8	Placed
M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.5	Mkt&Fin	66.28	Placed
M	65	Central	68	Central	Arts	64	Comm&M	No	75	Mkt&Fin	57.8	Placed
M	56	Central	52	Central	Science	52	Sci&Tech	No	66	Mkt&HR	59.43	Not Placed
M	85.8	Central	73.6	Central	Commerc	73.3	Comm&M	No	96.8	Mkt&Fin	55.5	Placed
M	55	Others	49.8	Others	Science	67.25	Sci&Tech	Yes	55	Mkt&Fin	51.58	Not Placed
F	46	Others	49.2	Others	Commerc	79	Comm&M	No	74.28	Mkt&Fin	53.29	Not Placed
M	82	Central	64	Central	Science	66	Sci&Tech	Yes	67	Mkt&Fin	62.14	Placed
M	73	Central	79	Central	Commerc	72	Comm&M	No	91.34	Mkt&Fin	61.29	Placed
M	58	Central	70	Central	Commerc	61	Comm&M	No	54	Mkt&Fin	52.21	Not Placed
M	58	Central	61	Central	Commerc	60	Comm&M	Yes	62	Mkt&HR	60.85	Placed
M	69.6	Central	68.4	Central	Commerc	78.3	Comm&M	Yes	60	Mkt&Fin	63.7	Placed
F	47	Central	55	Others	Science	65	Comm&M	No	62	Mkt&HR	65.04	Not Placed
F	77	Central	87	Central	Commerc	59	Comm&M	No	68	Mkt&Fin	68.63	Placed
M	62	Central	47	Central	Commerc	50	Comm&M	No	76	Mkt&HR	54.96	Not Placed
F	65	Central	75	Central	Commerc	69	Comm&M	Yes	72	Mkt&Fin	64.66	Placed
M	63	Central	66.2	Central	Commerc	65.6	Comm&M	Yes	60	Mkt&Fin	62.54	Placed

Fig. 2: Data set Uploaded in Weka

MODEL GENERATION

Models are generated using the given dataset of students placed at the campus. Following machine learning models are generated and tested using cross-validation technique using 10 fold.

DECISION TREE CLASSIFIER

The most frequently, and currently perhaps the most extensively used decision tree algorithm is C4.5. Professor Ross Quinlan developed a decision tree algorithm known as C4.5 in 1993. It characterizes the outcome of research that traces back to the ID3 algorithm which is also proposed by Ross Quinlan in 1986. C4.5 has extra features such as handling missing values, classification of continuous attributes, trimming of decision trees, rule derivation, and others. J48 algorithm is an implementation of the C4.5 decision tree algorithm in the Weka software tool.

ISSN:2395-1079
Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619
South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

The generated decision tree model is shown in Fig 3.

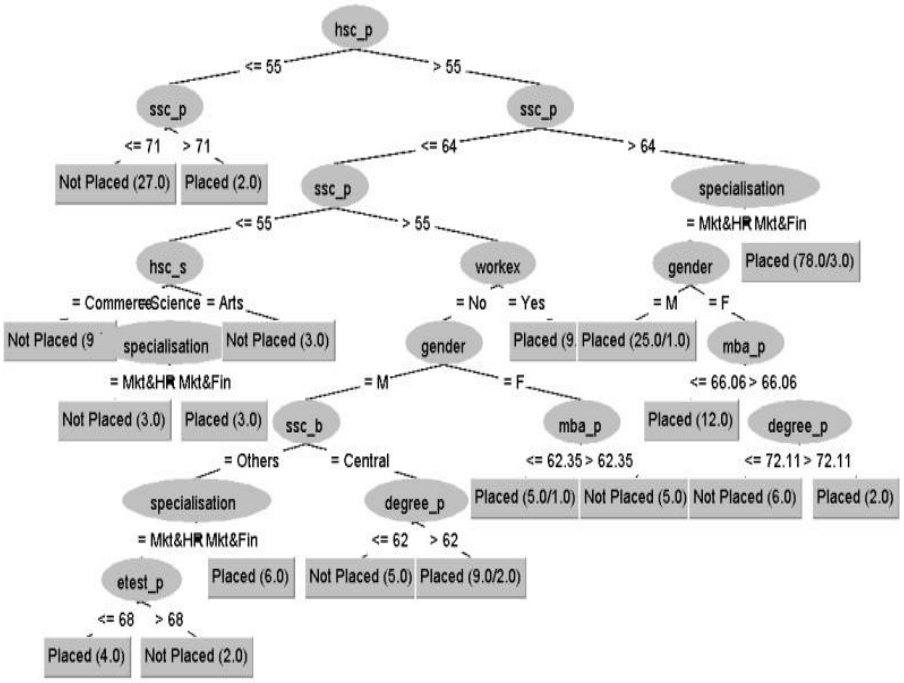

Fig. 3: Decision Tree

The confusion matrix generated by testing data using a decision tree algorithm is given in Fig 4. The accuracy achieved by this model is 82.79 % in which 178 instances are correctly classified.

```
=== Confusion Matrix ===

   a   b   <-- classified as
 131  17 |   a = Placed
  20  47 |   b = Not Placed
```

Fig. 4: Decision Tree Confusion Matrix

NAÏVE BAYES

The Naïve Bayes model detects the features of failure students. It illustrates the probability of each input feature for the predictable state. A Naive Bayesian classifier is a simple probabilistic classifier based on applying the Bayesian theorem with robust independence

ISSN:2395-1079
Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619
South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

suppositions. Naive Bayes employment is chosen as it is simple and can be trained on whole training data, a privilege that boosting "never over-fits" could not be preserved, and the complex resultant classifier can be determined constantly the naïve bayes classifier is shown in Fig 5. from a limited amount of data [8]. The accuracy achieved by classifying the dataset using naïve bayes classifier is 84.65% in which 182 out of 215 instances are correctly classified. The confusion matrix of

```
=== Confusion Matrix ===

   a   b   <-- classified as
 132  16 |   a = Placed
  17  50 |   b = Not Placed
```

Fig. 5: Confusion Matrix of Naive Bayes

RANDOM FOREST

The random forest algorithm can also be thought of as an ensemble method in machine learning. The input to a random forest algorithm is a dataset consisting of records, with attributes. Random subsets of the input are created. On each of the random subsets created, a decision tree will be constructed. The final class of a test record will be decided by the algorithm which uses the majority vote technique. Random forest algorithm makes use of the out-of-bag error technique [9]. The accuracy obtained by using a random forest classifier is 86% in which 185 instances are correctly classified.

```
=== Confusion Matrix ===

   a   b   <-- classified as
 140   8 |   a = Placed
  22  45 |   b = Not Placed
```

The confusion matrix generated by the Weka tool using random forest is given in Fig. 6.

Fig. 6: Confusion Matrix of Random Forest

ISSN:2395-1079
Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619
South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

RESULTS ANALYSIS

To better understand the importance of the input variables, it is customary to analyze the impact of input variables during students' placement success, in which the impact of certain input variables of the model on the output variable has been analyzed. Different algorithms provide very different results, i.e. each of them accounts for the relevance of variables in a different way. Now, we have carried out some experiments to evaluate the performance and usefulness of different classification algorithms for predicting students' placement. The results of the experiments are shown in table 1.

Table 1: Performance of Classifiers

Evaluation Criteria	Classifiers		
	Decision Tree	Naïve Bayes	Random Forest
Time to build model	0.01 seconds	0.001 seconds	0.1 seconds
Correctly Classified Instances	178	182	185
Incorrectly Classified	37	33	30
Instances			
Accuracy (%)	82.79 %	84.65 %	86.04 %
Recall	0.885	0.892	0.946
Precision	0.868	0.886	0.864

The efficiency of the three approaches is compared in terms of accuracy, recall, and precision. The accuracy of the prediction model is defined as the total number of correctly predicted/classified instances. Accuracy is given by using the following formula:

$$\text{Accuracy} = (TP + TN)/(TP+FP+TN+FN) * 100$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and falsenegatives.

Precision also called positive predictive value is the fraction of relevant instances

ISSN:2395-1079
Available online at https://journals.edwin.co.in/index.php/esajms/issue/view/619
South Asia Journal of Multidisciplinary Studies SAJMS June 2022, Vol. 8, No.-5

Research Paper

among the retrieved instances, while recall is also known as sensitivity is the fraction of relevant instances that were retrieved. Both precision and recall are therefore based on relevance.

We had done the comparative analysis of all three models, namely, Decision Tree, Random Forest, and Naïve Bayes. Fig. 7 shows the graph of the comparison of the models taken under consideration. These models were selected for our research work as these are the best-suited models for classification problems.

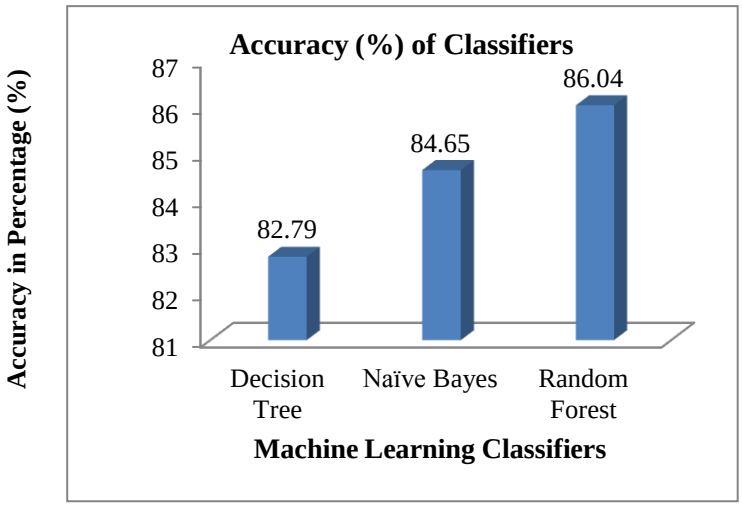

Fig. 7: Classifiers Accuracy

CONCLUSION

The campus placement task is extremely a lot of vital from the organization's point of view as well as the student's point of view. In this respect to advance the student's performance, an effort has been studied and predicted using the classification algorithms Decision Tree, Naïve Bayes, and the Random forest algorithm to authenticate the methodologies. The algorithms are applied to the data set and features are selected to build the model. The accuracy obtained after analysis for the Decision tree is 83%, for Naïve Bayes is 84.65% and for the Random Forest is 86%. These results recommend that amongst the machine learning algorithm verified, the Random Forest classifier has the potential to significantly progress the conventional classification methods for use in placement.

Research Paper

REFERENCES

[1] https://www.kaggle.com/benroshan/factors-affecting-campus-placement
[2] Manvitha, Pothuganti, and Neelam Swaroopa. "Campus placement prediction using supervised machine learning techniques." Int J Appl Eng Res 14.9 (2019): 2188-2191.
[3] S. K. Thangavel, P. D. Bkaratki and A. Sankar, "Student placement analyzer: A recommendation system using machine learning," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), 2017, pp. 1-5, doi: 10.1109/ICACCS.2017.8014632.
[4] Laxmi Shanker Maurya and Md Shadab Hussain and Sarita Singh, "Developing Classifiers through Machine Learning Algorithms for Student Placement Prediction Based on Academic Performance", In Applied Artificial Intelligence, vol. 35, no. 6, pp. 403-420, 2021, doi: 10.1080/08839514.2021.1901032.
[5] Pratiwi, Oktariani Nurul. "Predicting student placement class using data mining." Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE). IEEE, 2013.
[6] Quinlan, J.R., C4.5: Programs for machine learning, Morgan Kaufmann, San Francisco, 1993.
[7] Kajal Rai, M. Syamala Devi, and Ajay Guleria, "Decision Tree based Algorithm for Intrusion Detection", International Journal of Advanced Networking Applications (IJANA), ISSN: 0975-0282, vol. 7, no. 4, pp. 2828-2834, 2016.
[8] Russell, Stuart, Norvig, Peter, "Artificial Intelligence: A Modern Approach", In Prentice Hall, (1995), ISBN: 978-0137903955.
[9] Livingston, Frederick. "Implementation of Breiman's random forest machine learning algorithm." ECE591Q Machine Learning Journal Paper (2005): 1-13.