Journal of Advanced Zoology
ISSN: 0253-7214
Volume 45 Issue S-4 Year 2024 Page 06 -12

Maximizing Campus Placement Through Machine Learning

Ms. Sarita Byagar[1*], Dr. Ranjit Patil[2], Dr. Janardan Pawar[3]

[1]Research Scholar, Research Centre in Commerce and Management, Indira College of Commerce and Science, Affiliated to Savitribai Phule Pune University, and Assistant Professor, Department of Computer Science, Indira College of Commerce and Science, Pune-33.*
[2]Research Guide, Research Centre in Commerce and Management, Indira College of Commerce and Science, Affiliated to Savitribai Phule Pune University, and In-Charge Principal, Dr. D.Y. Patil Arts, Commerce & Science College. Pimpri.
[3]Principal In-charge, Indira College of Commerce and Science, Pune- 33.

***Corresponding Author:** Ms. Sarita Byagar*
**Research Scholar, Research Centre in Commerce and Management, Indira College of Commerce and Science, Affiliated to Savitribai Phule Pune University, and Assistant Professor, Department of Computer Science, Indira College of Commerce and Science, Pune-33.*

Abstract

Campus placement is an essential aspect of a student's academic career, as it determines their entry into the workforce. Predicting students' campus placement can help universities and colleges identify students who are likely to be successful in their chosen career paths and provide them with the necessary support to secure a job. Application of machine learning algorithms to forecast students' placement on campuses has gained traction in recent years. This research paper will explore the different approaches to predicting students' campus placement, the factors that influence campus placement, and the benefits and limitations of using machine learning algorithms for prediction. Machine learning is capable of adaptability and with the use of statistical models and algorithms they are able to draw inferences from patterns in data. Using ML algorithms forecasting can be done about the campus placement of students. Three ML algorithms viz, Naïve Bayes, Random Forest and Decision Trees are used to forecast the job/campus placement of students and evaluation of the aforesaid algorithms are performed with respect to accuracy of the classifier[11].

Keywords: Machine Learning, Prediction, Naïve Bayes, Random Forest, Decision Tree, Placement.

INTRODUCTION

Educational Institutions are increasing in high numbers and objective of every higher education's institution is to get their students placed in well-off companies with a high paid job. Campus placements have become the buzzword now. For any college and student, the placements hold exceptional significance. Before anyone takes admission to a college/institute the first focus is always on campus placements. Campus placement is an entry point for all the newbies in the corporate world and for this opportunity students give their best so that they

start their career at right time. Students can begin their careers precisely when their coursework is completed thanks to campus placements. Additionally, they get the opportunity to communicate and engage with the enterprise professionals, which helps to groundwork their future career path by acquainting them with knowledgeable people in their chosen field of expertise. [4]

a) Machine Learning Methods Utilization on Campus Placement

Machine Learning is a growing technology and as the name suggests makes the machine capable to learn and decide from past data and helps to make rational predictions and classifications. It is used for a variety of tasks viz. image processing, natural language processing, speech recognition, spam filtration etc. Machine learning can be done in two ways: supervised and unsupervised. However, various datasets and different contexts call for the application of alternative approaches. Supervised learning is a machine learning technique that involves training models with labelled data. Unsupervised learning is alternative machine learning method in which patterns are formed from the unlabeled input data. There is plethora of supervised and unsupervised algorithms available for classification and prediction of data. [5]

b) Factors Influencing Campus Placement

Several factors can influence students' campus placement. Some of these factors includes academic Records, internship experience, communication skills, personality traits and many more

Academic record indicates the skill and grading obtained in subjects of the studied curriculum. The student's academic record is a crucial factor in campus placement. A good academic record can increase the chances of being placed in top companies. Internship Experience is the exposure obtained by internee in organization after completing graduate study. IT provides students necessary skills and experience required to secure a job. Communication Skills are the abilities of the student for effective communication. This is essential for campus placement. Students who can communicate well are more likely to be placed in top companies. Personality Traits are inter and external qualities of individual required for effective execution of the work. It includes leadership skills, team management, and adaptability can influence campus placement. [1]

RESEARCH OBJECTIVES

The following are some possible research objectives for "Campus Placement Prediction using Supervised Machine Learning Algorithms":

- To determine which supervised machine learning algorithms are best at predicting campus placements using past data.
- To collect and preprocess data from various sources, including student performance data, recruiter feedback, and placement statistics, for use in training and testing the machine learning models.
- To develop and optimize machine learning models that accurately predict the likelihood of a student getting placed based on their academic performance, skills, and other relevant factors.
- To evaluate the performance of different machine learning models using various metrics, such as accuracy, precision, recall, F1 score and compare them to identify the best-performing model.
- To examine the factors that influence campus placements and identify actionable insights that can help students and recruiters improve their placement outcomes.

RELATED WORK

Predicting students' campus placement has been the focus of several studies in recent years. Researchers have used different approaches such as decision trees, random forests, and support vector machines to predict campus placement. For instance, Guleria & Sood (2015) used a Bayesian classification technique to predict the campus placement of students. They found that the support vector machine model was effective in identifying students who were likely to be placed.

In another study, Manvitha, Pothuganti, and Neelam Swaroopa. (2019) used supervised machine learning algorithms to predict campus placements. They found that the decision tree model was effective in identifying students who were likely to be placed in top companies. Other factors such as the student's academic record, internship experience, and communication skills were also found to influence campus placement.

Based on a student's academic achievement in their tenth and twelfth classes, graduation, and backlog to date, supervised machine learning classifiers can be used to predict their placement in the industry. Knowledge Discovery and Data Mining (KDD), a placement class approach utilising the classification technique, was employed by Pratiwi, Oktariani Nurul. Using identical data and attributes, the second trial produced the best

percentage of accuracy at 92.1%. In the first experiment, 84.2% of the instances were correctly identified. The combination of SMO and Naive Bayes produced the greatest results.

PROPOSED METHODOLOGY

Methodology for "Campus Placement Prediction using Supervised Machine Learning Algorithms" is executed as follows:

a) ***Data Collection:*** Collecting relevant data related to past campus placements from multiple sources such as placement reports, student resumes, and recruitment data.

b) ***Data Preprocessing***: Cleaning and preprocessing the collected data by removing inconsistencies, missing values, and irrelevant features.

c) ***Feature Selection***: Selecting the most significant features using techniques such as correlation analysis and feature importance ranking.

d) ***Data Partitioning***: Splitting the preprocessed data into training and testing sets in a ratio of 70:30 or 80:20, respectively.

e) ***Algorithm Selection:*** Choosing appropriate supervised machine learning algorithms based on the problem statement, available data, and performance metrics.
In this research, the researchers have used three ***supervised learning algorithms*** *viz. Naïve Baise, Random Forest and Decision Trees, so-as-to predict the campus placement of students.*

f) ***Model Training***: Utilizing the training data to train the chosen algorithms and assessing their results using a variety of metrics, including accuracy, precision, recall, and F1-score.

g) ***Hyperparameter Tuning:*** Fine-tuning the hyperparameters of the chosen algorithms to improve their performance on the testing data.

h) ***Model Selection***: Comparing the performance of different algorithms and selecting the best-performing model based on the chosen evaluation metric.

The proposed methodology is depicted in the Fig.1.

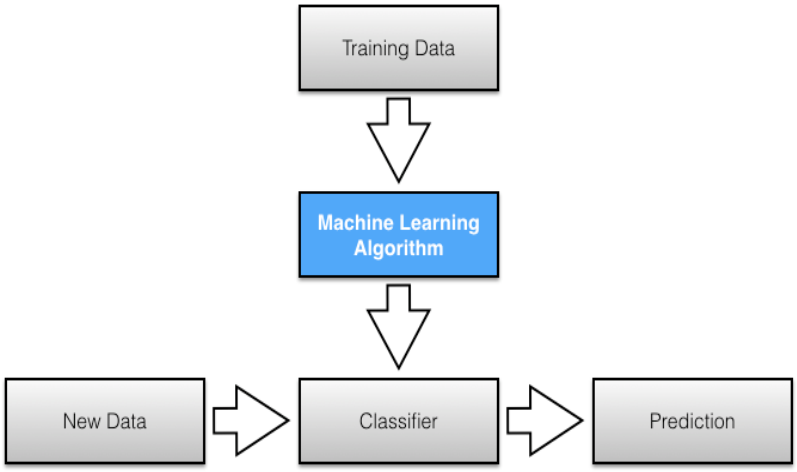

Fig.1: Proposed Methodology of Campus placement prediction [Compiled by Researcher]

EMPIRICAL WORK

a) Data Collection and Preprocessing

The sample data has been collected from IT courses viz. Computer Science, Computer Application students of INDIRA college in Pune. The dataset comprises of 215 instances of students. Preprocessing is required to make the data ready for analysis. There are quite a few steps involved in data preprocessing which includes data cleaning, handling missing values, and attribute selection. In a dataset, there can be few attributes which can may be irrelevant and hence hamper the accuracy of the result. Considering the same, the attributes which directly affect the classification and prediction are retained. In some cases, the dataset may contain blank or missing values which needs to be handled and equipped and with an appropriate value. A missing value can be replaced by a default value, or the mean of that column or most simple solution can be to remove the whole row.

b) Algorithm Selection

i. Naïve Bayes

The supervised machine learning method Naïve Bayes, which is utilised for classification, is founded on the Bayes theorem. Its a probabilistic classifier which means it predicts based on the probability of a certain event. As it works on the principle of Bayes theorem, the formula for the same is as follows: P(A/B) = (P(B/A).P(A))/P(B). Naive Bayes classifiers assume that value of a particular feature is independent of the value of other feature, given class variable.

ii. Decision Trees

Decision trees algorithm is a significant machine learning technique which is frequently used for classification and regression problems. Decision trees have a sequence well defined set of rules which is used to classify patterns. As the name suggests, a tree has a graph like structure which comprises of branch and leaf nodes which makes it easily understandable and applicable. Decision trees are generally preferred for rational decision making and prediction of forthcoming circumstances based on historical data[9]

iii.Random Forest

Random forest (RF) algorithm is one of the most popular machine learning technique used for classification and regression problems. Forest, the term refers to multiple trees and more the number of trees the algorithm will be robust enough. Hence more number of trees ensure higher accuracy and problem solving and correct prediction ability. Random forest develops from decision trees. There are multiple decision trees and when a new example/case has to be classified, each decision tree delivers a classification for the provided input data. Random forest gathers the classification result and uses quorums/votes to choose the most preferred prediction as the result.

c) Training and Test Data

A set of data is called training data when it is used to train a model, and a piece of data is called test data when it is tested after training successfully. The next stage will be to divide our dataset into two after preprocessing. A test set and a training set. First, we will use our training set to train our machine learning models, which will attempt to identify any connections in the data. Next, we will use our test set to evaluate the models' predictive accuracy. Assigning 80% of the dataset to the training set and the remaining 20% to the test set is a common practice.

d) Tool used for Experiment

The widespread tool WEKA was used for carrying out the analysis and prediction of the dataset. In the supervised learning category, there are various algorithms used. In this research, the algorithms used are Naïve Bayes, Decision Tree, and Random Forest algorithms. The dataset was analyzed for the aforesaid algorithms. For every algorithm, accuracy is the most important parameter which specifies how correctly the algorithm has classified the instances of the dataset. Apart from accuracy, Precision, Recall and F-score is also considered for the comparison of all three algorithms.

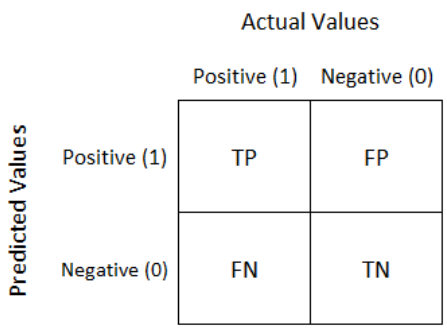

Fig. 2: Confusion Matrix To Be Used To Capture Empirical Results [9]

A table used to describe how well a classification algorithm performs is called a confusion matrix. A confusion matrix visualizes and summarizes the performance of a classification algorithm [9]. This matrix consists of True positive (TP): Observation is predicted positive and is positive. False positive (FP): Observation is

predicted as positive and is negative. True negative (TN): Observation is predicted negative and is negative. False negative (FN): Observation is predicted negative and is actually positive.[9][11] .
The number of positive class predictions that are part of the positive class is quantified by precision. [9] The amount of accurate class predictions made from all of the dataset's positive examples is measured by recall. [9]. The precision and recall problems are combined into a single score using the F-Measure.[9]

RESULT ANALYSIS

a) Experiment Result

After executing the three mentioned algorithms, the results obtained are placed in Table-1, Table-2 and Table-3 respectively. Based on these tables, algorithms are evaluated using the metrics accuracy, Recall, Precision and F-Score, which is shown in Table-4.

Table-1: Confusion Matrix for Naïve Bayes Classifier

a	b	<-- classified as
131	17	a = Placed
16	51	b = Not Placed

Table-2: Confusion Matrix for Decision Tree Classifier

a	b	<-- classified as
145	3	a = Placed
27	40	b = Not Placed

Table-3: Confusion Matrix for Decision Tree Classifier

a	**b**	**<-- classified as**
148	0	a = Placed
0	67	b = Not Placed

Table-4: Performance analysis of algorithms.

Evaluation Parameters	**Naïve Bayes**	**Decision Tree**	**Random Forest**
Correctly Classified Instances	182	185	215
Incorrectly Classified Instances	33	30	0
Accuracy	84.65 %	86.04%	100%
Recall	0.891	0.980	1
Precision	0.885	0.843	1
F Score	0.888	0.906	1

As per the results obtained, Random Forest gives best result..

b) Graphical Analysis

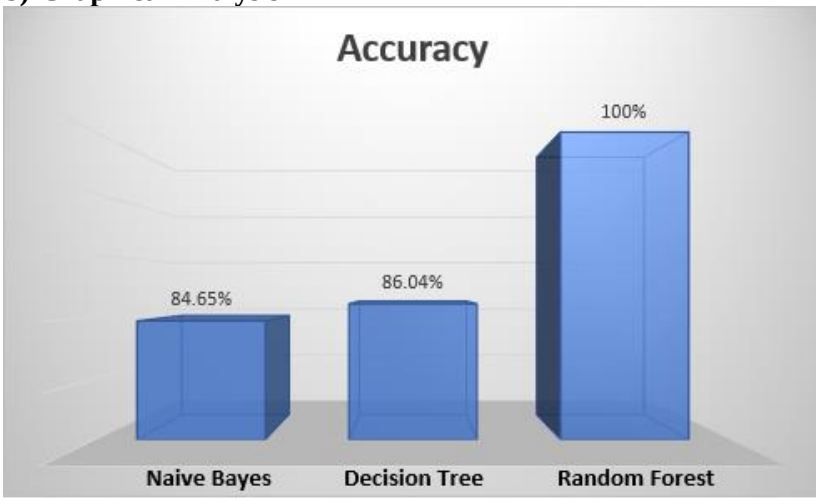

Fig.3: Performance of Classifiers [Compiled by Researcher]

c) Discussion

Predicting students' campus placement is essential for universities and colleges to identify students who are likely to be successful in their chosen career paths. Machine learning algorithms have the potential to improve the accuracy of predicting students' campus placement. In this respect an effort to study and predict the campus placements using the supervised machine learning classification algorithms Decision Tree, Naïve Bayes, and the Random forest algorithm have been used in this study. The algorithms are applied to the data set and features are selected to build the model. As seen in Table 4 and Figure 3, the accuracy obtained after analysis for the Decision tree is 86.04%, for Naïve Bayes is 84.65% and for the Random Forest is 100%. These findings suggest that the Random Forest classifier has the ability to more accurately classify and predict student placements among the several supervised machine learning methods.. This study can be definitely of great help to colleges and universities to carry out the placement process smoothly. However, there are several limitations to using these algorithms, such as lack of transparency and bias which needs to be taken care of.

CONCLUSION

Some common supervised machine learning algorithms used in campus placement prediction include logistic regression, decision trees, random forests, support vector machines, and artificial neural networks. These algorithms can be trained on historical campus placement data to learn the relationships between various factors such as academic performance, work experience, and personal characteristics, and the likelihood of getting placed.

The evaluation of the performance of these algorithms is typically done using metrics such as accuracy, precision, recall, F1-score. The results of these evaluations can help in selecting the most effective algorithm for a given dataset and application scenario.

In conclusion, supervised machine learning algorithms have the potential to improve the accuracy and efficiency of campus placement prediction. However, it is important to carefully choose and evaluate the algorithms, as well as the dataset and feature selection process, to ensure the best possible results.

FUTURE RESEARCH WORK

There are several potential areas for future research work on campus placement prediction using supervised machine learning algorithms. Some possible directions are:

- ***Feature selection and engineering:*** Although many studies have focused on selecting the most relevant features for campus placement prediction, there is still room for improvement in this area. Future research could explore more sophisticated techniques for feature selection and engineering, such as deep learning-based methods or natural language processing techniques to extract features from resumes.
- ***Ensemble learning:*** Ensemble learning is a technique that combines multiple models to improve prediction accuracy. Future research could investigate the use of ensemble learning techniques for campus placement prediction, such as stacking or boosting methods.
- ***Fairness and bias:*** Campus placement prediction models may inadvertently incorporate biases, such as gender or race-based discrimination, which can lead to unfair outcomes. Future research could explore techniques for mitigating these biases, such as algorithmic fairness measures or pre-processing methods to ensure fairness.
- ***Online and real-time prediction:*** Many campus placement prediction models are designed to be used offline, based on historical data. Future research could focus on developing online and real-time prediction models that can be used to predict placement outcomes in real-time, based on streaming data from job applicants.
- ***Comparative analysis of algorithms:*** Although several supervised machine learning algorithms have been applied to campus placement prediction, there is a need for comparative analysis to determine the most effective algorithm for a given dataset and application scenario. Future research could focus on comparative analysis of different algorithms to identify the strengths and weaknesses of each algorithm in terms of prediction accuracy and efficiency.

REFERENCES

1. Shreyas Harinath, Aksha Prasad, Suma H and Suraksha A. Student Placement Prediction using Machine Learning, International Research Journal of Engineering and Technology (IRJIET) Volume: 06 Issue: 04 April 2019
2. Senthil Kumar Thangavel, Divya Bharathi P and Abhijith Shankar. Student Placement Analyzer: A recommendation System Using Machine Learning, International Conference on advanced computing and Communication systems (ICACCS-2017), Jan 06- 07,2017, Coimbatore, INDIA.
3. K. Sreenivasa Rao, N. Swapna, P. Praveen Kumar Educational data mining for student placement prediction using machine Learning algorithms Research Paper, International Research Journal of Engineering and Technology (IRJIET) 2018
4. Manvitha, Pothuganti, and Neelam Swaroopa. "Campus placement prediction using supervised machine learning techniques." Int J Appl Eng Res 14.9 (2019): 2188- 2191.
5. 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.
6. 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.
7. 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
8. Guleria, P., & Sood, M. (2015, December). Predicting student placements using Bayesian classification. In 2015 Third International Conference on Image Information Processing (ICIIP) (pp. 109-112). IEEE.
9. www.sciencedirect.com
10. https://www.wikipedia.org
11. Byagar, S., & Thakare, S. Harnessing the power of machine learning for predicting students employability.