

RESEARCH

Prognostication of the placement of students applying machine learning algorithms

Gladrene Sheena Basil*

Department of Computer Science and Engineering, Loyola Institute of Technology and Science, Nagercoil, India

***Correspondence:**Gladrene Sheena Basil,
gladrene.basil@gmail.com**Received:** 26 May 2023; **Accepted:** 07 July 2023; **Published:** 20 September 2023

Placement is the process of connecting the selected candidate with the employer. Every student might have a dream of having a job offer when he or she is about to complete her course. All educational institutions aim at having their students well placed in good organizations. The reputation of any institution depends on the placement of its students. Hence, many institutions try hard to have a good placement cell. Classification using machine learning may be utilized to retrieve data from the student-databases. A prediction model that can foretell the eligibility of the students based on their academic and extracurricular achievements is proposed. Related data was collected from many institutions for which the placement-prediction is made. This paradigm is being weighed up with the existing algorithms, and findings have been made regarding the accuracy of predictions. It was found that the proposed algorithm performed significantly better and yielded good results.

Keywords: placement, prediction, logistic regression, K-NN

Introduction

Every institution has its own placement cell and it contributes a decisive role toward the reputation of the institution. The success of any institution is measured by where the students of the institution are placed. The students make admission to a college by noticing the percentage of placements in the college. Therefore, a prediction model is necessary to analyze the placement and find out how many students are likely to get placed immediately. This will help to build the remaining students to become eligible and improve their placement opportunities. The proposed model will help in predicting whether a student will get placement or not. It will also be helpful for identifying the areas where students need to work so as to get a placement by the time they finish their course. Academic marks, achievements, regularity in attending classes, attendance, coding skills, team-playing ability, soft skills and extracurricular activities are taken into account. The placement statistics of the previous year and the student dataset are taken for the placement prediction. Thus, students may have an opportunity to make themselves readily employable.

Advantages in using machine learning

Machine learning focuses on using algorithms and data so that the way humans learn is exactly imitated and the accuracy of the algorithm is gradually improved. The types of machine learning are shown in [Figure 1](#).

Machine learning happens in exactly the same way that a person teaches a child. The computer learns to work by imitating the mechanisms of the human brain. The neural network is the basis of machine learning and it is designed in the same manner as the work of the human brain with neurons and dendrons. An outline of how machine learning works is given in [Figure 2](#).

Proposed scheme

This section outlines the research methodology espoused here. The acquisition of data is the first step. This step is followed by enhancement of data which prepares the data

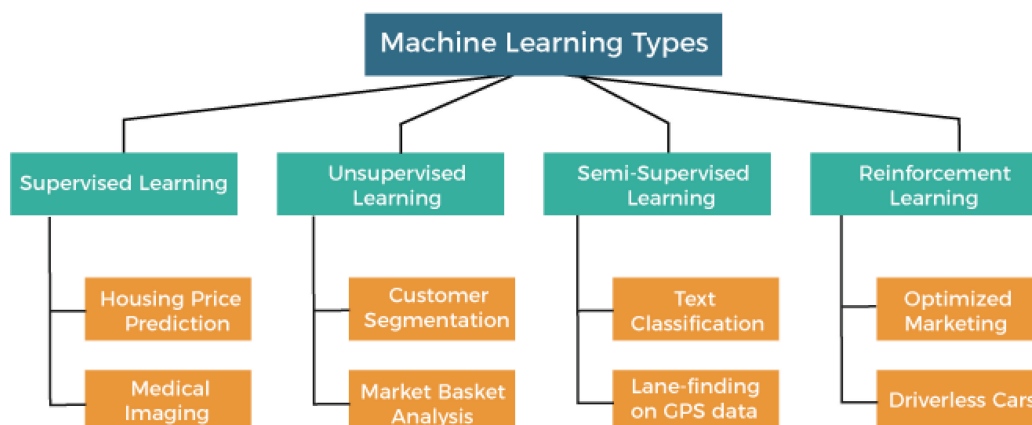


FIGURE 1 | Categories in machine learning.

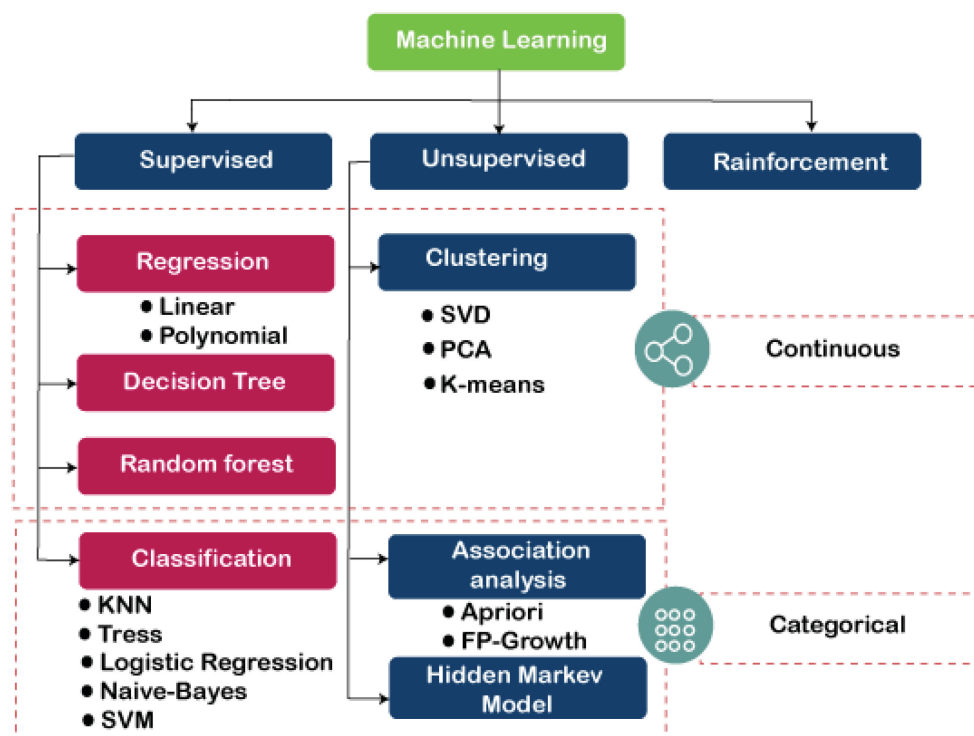


FIGURE 2 | Subdivisions in each of the categories in machine learning

for the next process. This is very important since this is where the data is made suitable for further processing. Data cleaning, data transformation and data reduction happen during pre-processing. Following the pre-processing, the actual processing happens. Finally, the last step is the interpretation of the data shown in **Figure 3**.

Acquisition of data

An enormous amount of data is required for the machine learning models to operate on and create the best paradigm for prediction. Hence, the dataset should be big enough and we can choose our data set from neighboring geographical

areas. The sample data was collected from the placement department. It consists of all the records of the students of previous years. The dataset belongs to over 1000 students. The dataset involves a number of events of each student consisting of cumulative grade point average (CGPA), other achievements, both academic and non-academic and extracurricular activities.

Pre-processing of data

When data has been acquired successfully, the next step is pre-processing. The data, which is collected from various sources, is in a raw form that is not feasible for any direct

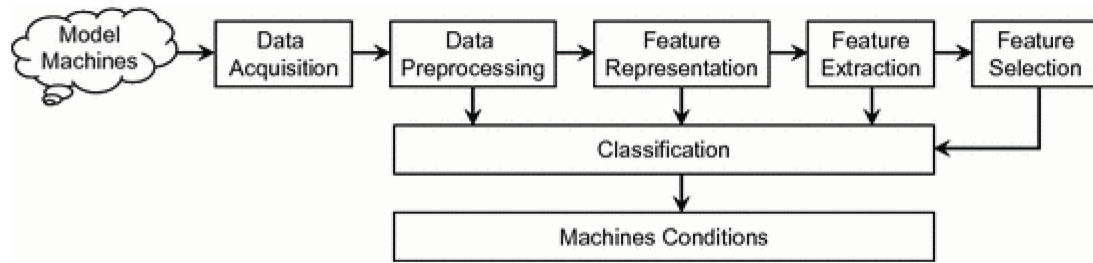


FIGURE 3 | Steps in pre-processing of data

analysis. The preprocessing of data consists of three steps. They are (i) Data cleansing (cleaning), (ii) Data alteration (transformation), and (iii) Data lessening (reduction). Only after these steps are over, the data passed on to the next stage. Pre-processing data is a procedure whose purpose is the conversion of raw data into a clean dataset.

Feature representation

Feature representation is the encoding process that maps the raw details onto a discriminant feature space. Feature representation is a class of procedures that help a system discover the representations needed for feature detection automatically from raw data.

Feature extraction

By means of feature extraction, new features that may not be actually present in the given group of features can be obtained to good advantage. The majority of sharp characteristics in signals are identified through the process of feature extraction. This facilitates easier consumption by machine learning or deep learning algorithms. The number of expedients that are needed to describe an enormous data set is diminished by feature extraction. This step starts with a primary group of assessed data and constructs secondary values that are meant to be didactic and necessary.

Feature selection

Feature selection is the operation of picking a subcategory of features that impart the maximum. Feature Selection is an approach in which the input variable to the model is diminished, using only relevant data and removing the noise in the data. It is an action in which relevant features for your machine learning model are chosen automatically based on the category of issue that you are attempting to work out. Specific information is collected from the institutions. The primary details taken for this study are name, stream, orientation, academic records, CGPA, extracurricular activities, and other achievements. Feature

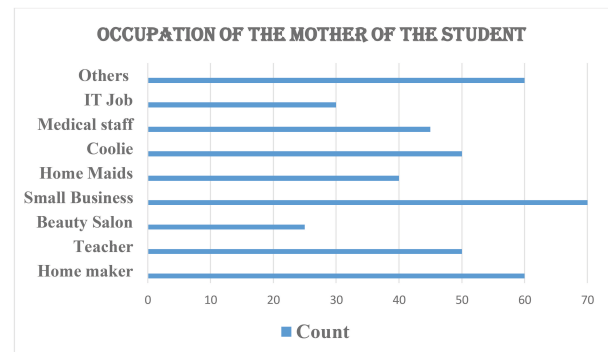


FIGURE 4 | Occupation of the mother of the student.

selection in machine learning is analogous to the selection of variables, a subset of variables, or an attribute. It is the operation of selecting a feature subset that may be applicable to the construction of the paradigm.

Classification

Classification, or categorization, is a machine learning method that is supervised. In this step, the model seeks to predict the correct label of a specified input data set. The model is entirely trained in this step using the training data. Then, it is judged on test data prior to being applied to perform predictions on unobserved and new data. Categorization is used to distinguish the class of new observations based on the training data.

Training and test data

The dataset is split into two subdivisions. One is called training data, which is a section of our original dataset that is supplied into the machine learning paradigm to learn patterns. Through this method, our model is trained. When the machine learning model is constructed with the training data, the unobserved data is required to examine the model. This is the testing data that can be utilized to assess the execution and advancement of the training in algorithms and improve it to get better results. The principal dissimilarity among training data and testing data is that training data is a

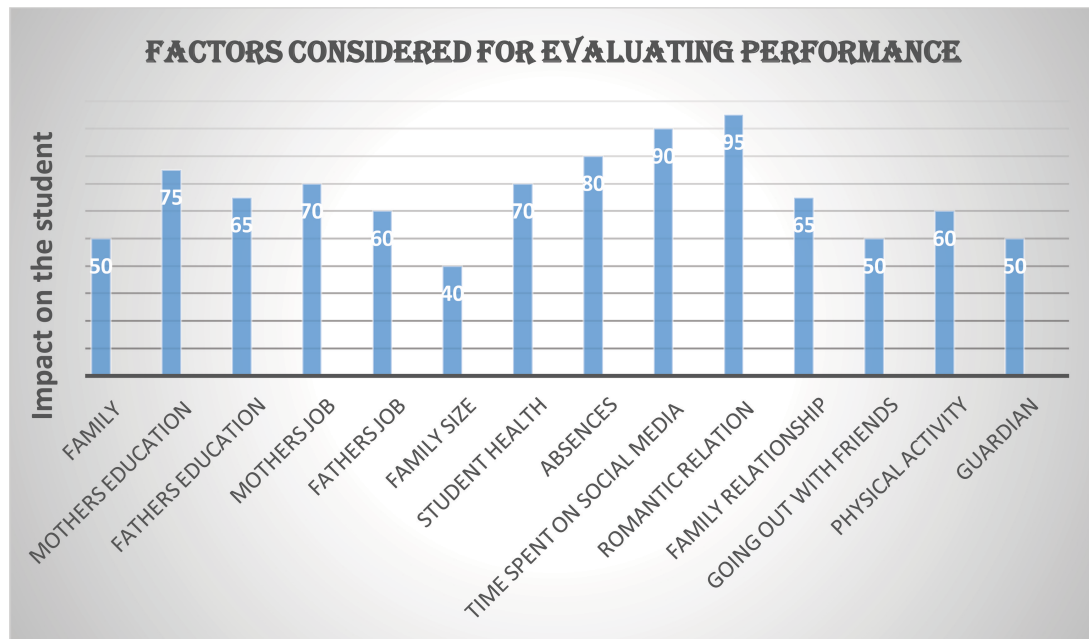


FIGURE 5 | Factors affecting the performance of the students.

subclass of the primary data that is used to train the machine learning model, while testing data is applied to substantiate the fidelity of the model. The training dataset is commonly larger in volume as compared with the testing dataset. The machine learning models will try to recognize and figure out the associations in the training set. Then testing is performed on the model. How the model makes the predictions is studied, and the accuracy is tested. Generally, a majority of the data is used for training, and a smaller set of the data is utilized for testing.

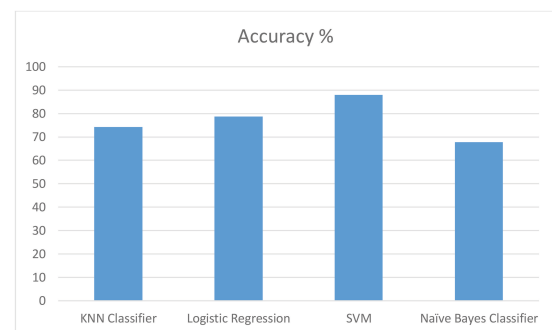


FIGURE 6 | Accuracy in percentage.

Categorization and prognosis

The data needs to be categorized (4) into classes or categories. This categorization is done in this paper by means of four supervised machine learning algorithms (5): support vector machine, KNN classifier, Naïve Bayes classifier, and logistic regression. After the algorithms were applied, the data were classified or categorized (6), and results were projected. This is an important step that helps in prognosis or prediction. Prediction will help in altering the situations for the better performance (7, 8, 9) of the students toward getting better placement opportunities (10, 11), as shown in [Figures 4, 5](#).

KNN classifier

KNN is a supervised algorithm that is applied to problems that are established by regression or classification. The KNN algorithm is employed to categorize the students (1) in either one of the categories, pass or fail. It is checked further to see whether it is functional. This technique reserves the

data, and each time a classification is made according to similarity in features.

Logistic regression

Logistic regression is also used for categorization. It is primarily used for forecasting by making use of independent variables and explicit dependent variables. The output of logistic regression is a value of probability between 0 and 1.

Applying the support vector machine

Support vector machine, or SVM, is among the top prevailing machine learning algorithms that are supervised. Both regression problems and classification problems make use of SVM. In the proposed method, the SVM is applied to categorize the students into either of the two categories, namely, employable or not. This technique constructs a

model that places new examples in one class. SVM plots in such a way as to increase the accuracy of classification.

Applying naïve Bayes classifier

The Naïve Bayes algorithm is a supervised one based on the Bayes theorem, and it is applied to working out classification puzzles. It is a straightforward method to build classifiers (2) that are nothing but models that give class labels that are picked from a definite set. A Naïve Bayes classifier computes the probability (3) of a class when a set of feature values is given.

Results and discussion

After the analysis was carried out, it was found that SVM was much superior to Naïve Bayes classifier, KNN classifier, and logistic regression. When the SVM was applied, 87.89% accuracy was achieved; the KNN classifier yielded 74.2% accuracy; logistic regression yielded an accuracy of 78.77%; and when the Naïve Bayes classifier was applied to the dataset, 67.85% accuracy was achieved. The comparison of the accuracy of these four models is shown in [Figure 6](#).

Conclusion

Thus, it is known that the SVM is much superior to the Naïve Bayes classifier, the KNN classifier, and logistic regression. The efficiency of the suggested model is clearly seen here. With the evidence shown, it is known that the proposed model to analyze students' performance will be effective and will give more insight on what affects the placements of the students. Better monitoring systems may be suggested, and educational institutions may introduce measures in order to assign grades scrupulously, thus improving the performance

of the students, making them more employable, and giving them bright placement opportunities. This prediction will help empower the students and equip them by helping them acquire more skills, both soft and technical. This model will definitely help to prepare the students and strengthen the recruitment training procedures so that a high percentage of placement is achieved.

References

1. Siva Surya M, Sathish Kumar M, Ganthimathi D. *Student Placement Prediction Using Supervised Machine Learning*. New York City, NY: IEEE explore (2022).
2. Shahane P. *Campus Placements Prediction and Analysis using Machine Learning*. New York City, NY: IEEE explore (2022).
3. Manike M, Singh P, Sai Madala P, Varghese SA, Sumalatha S. *Student Placement Chance Prediction Model using Machine Learning Techniques*. New York City, NY: IEEE explore (2021).
4. Maurya LS, Shadab Hussain M, Singh S. Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. *Appl Artif Intell*. (2021) 6:403–20.
5. Kumar N, Shanker Singh A, Thirunavukkarasu K, Rajesh E. *Campus Placement Predictive Analysis using Machine Learning*. New York City, NY: IEEE explore (2020).
6. Viram R, Sinha S, Tayde B, Shinde A. Placement prediction system using machine learning. *Int J Creat Res Thoughts*. (2020) 8:4.
7. Kumar N, Thirunavukkarasu K, Singh AS. Campus Placement Predictive Analysis using Machine Learning. *Proceedings of the 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*. Greater Noida (2020).
8. Manvitha P, Swaroopa N. Campus placement prediction using supervised machine learning techniques. *Int J Appl Eng Res*. (2019) 14:9.
9. Yang L. Application of machine learning techniques in college students' information system. *Proceedings of the International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2018)*. Amsterdam (2018).
10. Ishizue R, Sakamoto K, Washizaki H, Fukazawa Y. Student circumstance and capacity situating markers for programming classes using class attitude, mental scales, and code estimations. *Res Pract Technol Enhanc Learn*. (2018) 13:7.
11. Giri A, Vignesh M, Bhagavath V, Pruthvi B, Dubey N. *A Placement Prediction System using k-Nearest Neighbours Classifier*. New York City, NY: IEEE Explore (2016).