

AN EFFICIENT MACHINE LEARNING APPROACH FOR ACADEMIC PERFORMANCE AND PLACEMENT PREDICTION

*A. RAVI SANKAR, *Associate Professor and HOD, Department of CSE,*
SRINIVASA INSTITUTE OF TECHNOLOGY AND SCIENCES, KADAPA, AP.

ABSTRACT: Colleges and institutions allocate substantial resources to the placement process. The placement variables significantly influence the naming of institutions and the individuals who opt to attend. Consequently, educational institutions such as colleges and universities allocate substantial resources toward enhancing their programs to facilitate student employment. The objective of this paper is to evaluate student placement statistics from the preceding academic year, formulate informed projections regarding current student placement outcomes, and propose measures to enhance the placement rate within the educational system. This paper presents a methodology to assist educational institutions in selecting candidates for admission. When a student is correctly positioned, they can utilize historical data from prior students employed by the same organization to forecast the company's future performance. Machine learning employs a range of classification algorithms, such as the Naive Bayes Classifier and the K-Nearest Neighbors (KNN) model. These two algorithms categorize materials into specific categories. The dataset is utilized to evaluate the performance of the algorithms, which subsequently predict future events. The company's placement department may utilize the aforementioned framework to identify prospective candidates and assist them in concentrating their efforts on enhancing their technical and interpersonal skills.

Key Words: *Machine Learning, Naive Bayes, K-Nearest Neighbors (KNN), Database.*

1. INTRODUCTION

The quantity of newly established institutions has significantly risen in recent years. The primary function of the placement office, present in nearly all universities, is to assist students in securing lucrative employment opportunities. Colleges and universities face significant pressure to enhance the employment rates of their alumni. The caliber of the available educational alternatives is directly related to the intricacy of the placement. A comprehensive quality management system must incorporate the monitoring of evolving educational techniques and institutions. Data is systematically collected from diverse sources, including research databases, archives, and expert knowledge repositories, employing machine learning methodologies. Upon initial system startup, it utilizes a dataset of historical student data.

These results are utilized to evaluate the model's ability to effectively classify data and to assist in training the model to identify patterns and generalizations. Students exhibit increased motivation to excel academically when they recognize the abundance of employment prospects in various technology-related sectors. Educators utilizing this technology to

enhance student academic performance, together with the university's placement office, both benefit from it. A key indicator of a school's credibility is the proportion of graduates who secure employment post-graduation. Consequently, it is essential that this strategy be implemented in all university courses. We employ various machine learning methodologies, including K-Nearest Neighbors (KNN) and Naive Bayes, to achieve this objective.

PREDICTION SYSTEM

In order to arrive at placement decisions that are well-informed and based on textual input, the purpose of this work is to make use of machine learning techniques. Techniques from the field of machine learning, such as Naive Bayes and the K-Nearest Neighbors (KNN) algorithm, are applied in order to ascertain the precise location of the object. In addition to taking into account other relevant aspects, this method takes into account a candidate's technical and interpersonal skills as well as other relevant aspects. There are a number of factors that should be taken into consideration, including a person's cumulative grade point average (CGPA), their ranking in the United States Navy, their achievements in the tenth grade, and whether or not they have completed the Pre-University Course (PUC) or got their diploma.

NAIVE BAYES CLASSIFIER

The Naive Bayes Classifier has demonstrated its utility in numerous real-world applications. The Naive Bayes methodology generally outperforms competing strategies when features are judiciously selected and univariate conditional probabilities are accurately calculated.

K-nearest neighbor (KNN)

K-nearest neighbors (KNN) is a technique that deduces the characteristics of new data points by utilizing the entire training dataset. The k examples exhibiting the greatest similarity are selected following a meticulous analysis of the training set data. The data exhibiting the greatest degree of similarity is subsequently collected and shown.

Features of KNN

The entire training dataset is retained when employing the K-Nearest Neighbors (KNN) method.

The initial stage in generating real-time predictions is determining the similarity of each training example to an input sample.

2. RELATED WORK

Senthil Kumar Thangavel, Divya Bharathi P, and Abhijith Shankar focused their research on the development of placement prediction systems. The grade point average and place of residence of the individual were considered. The researchers employed Decision Tree Learning and SCI-Kit machine learning methodologies for the prediction analysis. Several variables, including the student's cumulative grade point average and any outstanding debts at the time of assessment, were considered in the projections.

In order to determine the most efficient method for students to select the classes that they will ultimately enroll in, a neural network was deployed in conjunction with colleagues Wilton Y.S. He, H.H. Au Yeung, K.Y. Law, and W.T. FOK. This enabled the researchers to determine the most effective technique. TensorFlow has undergone modifications and evaluations in order to ascertain the right amount of deep learning layers and intermediary nodes. These modifications and evaluations have been carried out.

Research in the discipline of machine learning entails the process of designing, testing, and analyzing algorithms that are capable of independently recognizing patterns in data and applying those patterns to forecast fresh data or inform wise choices. This process is known as "pattern recognition." In this particular scenario, machine learning helps to expedite the process by making it simpler to develop models that, when applied to circumstances that occur in the real world, produce outcomes that are satisfactory. Machine learning is gaining popularity in a wide range of domains, including bioinformatics, computer vision, robotic locomotion, computational finance, and search engine technology, among others. Its application is becoming increasingly pervasive.

3. METHODOLOGY

NAIVE BAYES

Classifiers, as the name implies, are software programs that assist computers in identifying the issue. These challenging circumstances are frequently shown using feature value vectors. The selection of class names is limited. Computer programs called Naive Bayes classifiers assist computers in distinguishing between groupings. One feature's value is independent of the values of other features for a given class variable. It's critical to understand that no single tool is all-inclusive. Among other reasons, apples are round, red, and 11 cm wide, which makes them excellent photographic subjects. The Naive Bayes classifier determines an object's likelihood of being an apple based on its color, shape, and thickness.

WORKING OF NAIVE BAYES ALGORITHM

An exhaustive analysis of all of the data that is currently saved on the computer is something that you are required to carry out before beginning any aspect of the process of data mining. Due to this, the only information that is retained is the one that is considered to be of the utmost significance. By making use of a variety of tools, including databases, cloud files, Excel spreadsheets, and other similar applications, you will be able to get this information in a relatively short period of time.

Before you can assess the likelihood of each potential combination of attribute numbers occurring, you must first compute the probability of that combination occurring. Within the scope of this particular investigation, the sections n , n_c , m , and p are going to be scrutinized. It is recommended that the following mathematical approach be performed in order to ascertain the probability that each attribute is accurate. This section provides a better and more detailed summary of the events that took place. In order to guarantee that the mathematical computations performed by everyone are accurate at all times, it is an absolute necessity that they remain constant.

When exactly does the third step begin to take place? Utilizing numerical data from a variety of different time periods throughout history By utilizing the method of $(n_c+mp)/(n+m)$, it is possible to ascertain the probability that the attribute value (a_i) will be present in the event that the subject value (v_j) is present. When it comes to this particular case, the variable n denotes the number of training scenarios in which the subject numbers v and v_j are equal to one another. All of these situations are considered to be identical. A further point to consider is that the value n_c is a representation of the frequency with which the subject value v and the characteristic value a correspond to values v_j and a_i , respectively. The fact that the value of

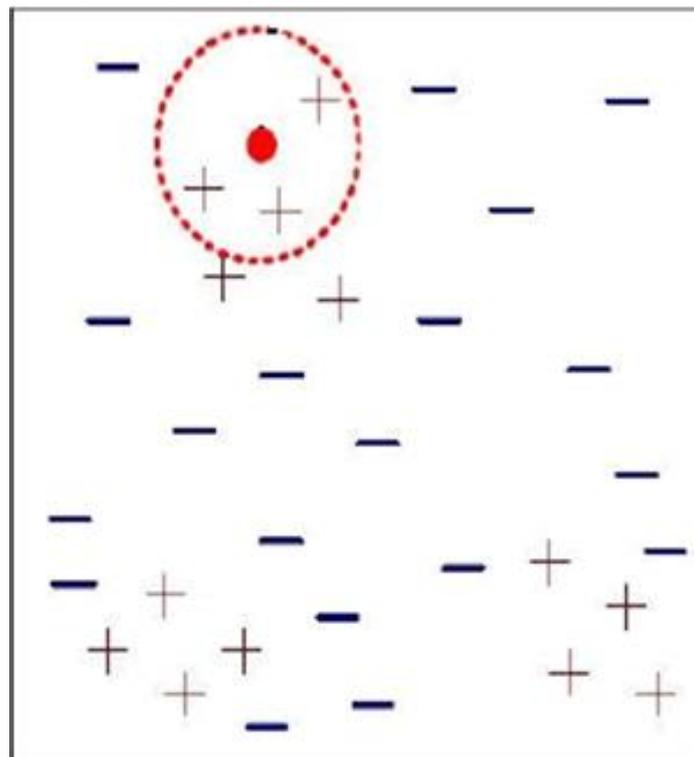
nc is bigger than zero is evidence that this is the case. Initially, the variable p is used to display a representation of the probability $P(a_{ij})$, which is shown at the beginning of the method.

It is essential to take note of the fact that the letter "M" stands for the total number of samples that were gathered.

To achieve additional outcomes, the fourth step includes increasing the probability value of each group by the entire number of possible outcomes, which is represented by the letter p . This is done in order to achieve additional outcomes. This approach makes use of a multiplier that is denoted by the letter p in order to make adjustments to the values that are associated with each attribute. In accordance with the conclusions of the study as a whole, the items have been divided into categories that are more precise. The primary objective is to arrive at a classification of each attribute number in accordance with the distinctive characteristics that it possesses.

K-nearest Neighbors classifier (KNN)

A straightforward method for grouping objects without the use of parameters is the K-Nearest Neighbors (KNN) method. One technique for using statistics to examine fixed traits or qualities is the K-nearest neighbors (KNN) method. The statistical technique known as K-nearest neighbors, or KNN, is not limited to pattern recognition and prediction. Making decisions about the present and the future by using the lessons you've learned from previous events is known as case-based reasoning, or CBR. Finding an unlabeled sample's closest friends is a quick and simple method of sorting it. Because it is easy to use and consistently produces the same results, the K-Nearest Neighbors (KNN) approach is frequently preferred over alternative classification methods. The arrangement of "near neighbors" is depicted in the image. There are two primary components to the K-nearest neighbors (KNN) approach.



Select the instance (in this case, S) in the group that most closely resembles the provided instance.

As previously stated, instances k use a political process to select the object type S . The number of nearest neighbors employed and the distance between neighbors determine how effectively the K -nearest neighbors (KNN) approach performs. Cosine and Euclidean distances are used by a large portion of the public to determine the separation between two occurrences.

Working of KNN Algorithm

Reports state that the proposed approach should improve the K -nearest neighbors (KNN) classifier's ability to evaluate health-related factors. The intended method's operation is demonstrated by the following formula.

Would you kindly open the file first?

- The K -Nearest Neighbors (KNN) approach is then applied to the data from this stage of the procedure.
- Determine how the measurements of the factors measured in the first phase and the third phase differed from another.
- Information regarding patterns is utilized at the conclusion of the process to assess the likelihood of the desired outcome.

The nearest neighbor algorithm measures the distance $d_E(X_i, X_j)$ between query points X_i and a set of training samples X_j to classify a new object based on majority of K -nearest neighbor category of Y attributes of training samples.

Query point $X_i = x_1, x_2, x_3, \dots, x_n$

Training Sample $X_j = x_1, x_2, x_3, \dots, x_n$

$$Dist(c_1, c_2) = \sqrt{\sum_{i=1}^N (attr_i(c_1) - attr_i(c_2))^2}$$

$$k - NearestNeighbors = \{k - MIN(Dist(c_i, c_{test}))\}$$

$$prediction_{test} = \frac{1}{k} \sum_{i=1}^k class_i \text{ (or } \frac{1}{k} \sum_{i=1}^k value_i)$$

4. FUTURE ENHANCEMENT

A key objective of the planned upgrade is to incorporate new components that will ultimately result in an increase in the precision of placement estimates. This will be accomplished by including new parts. By conducting regular analyses and evaluations of the potential ideas or suggestions that are being considered, it is possible to improve the operation of the system which is now being considered.

5. CONCLUSION

The Student Placement Predictor is an advanced technology that makes precise placement predictions through machine learning. The main areas of interest for education specialists are how to gauge and forecast children's academic performance. With the help of these forecasts, students can significantly increase their academic performance and grades. In the past, professionals used a limited set of criteria, like a student's grade point average (GPA) and delinquency history, to determine where to place them. This approach is believed to produce less accurate data. However, because the results of the other study forecast student placement

based on a greater variety of educational variables, they are more accurate.

REFERENCES

1. Maurya, L. S., Hussain, S., & Singh, S. (2021). Developing classifiers through machine learning algorithms for student placement prediction based on academic performance. *Applied Artificial Intelligence*, 35(6), 403–420.
2. Premalatha, N., & Sujatha, S. (2019). A comparative study on students placement performance using data mining algorithms. *International Journal of Scientific & Technology Research*, 8(10), 1806–1812.
3. Patel, N. K. M., Goutham, N. M., Inzamam, K. A., Kandi, S. V., & Vineet Sharan, V. R. (2022). Placement prediction and analysis using machine learning. *International Journal of Engineering Research & Technology (IJERT)*, ICEI – 2022 (Vol. 10, Issue 11).
4. Swaminarayan, P., & Ruparel, M. (2024). Student placement prediction using various machine learning techniques. *International Journal of Intelligent Systems and Applications in Engineering*, 12(3), 2107–2113.
5. Byagar, S., Patil, R., & Pawar, J. (2024). Maximizing campus placement through machine learning. *Journal of Advanced Zoology*, 45(S4).
6. Hariharan, V. J., Abdullah, A. S., Rithish, R., Prabakar, V., Selvakumar, S., & Suguna, M. (2022). Predicting students' placement prospects using machine learning techniques. *SSRG International Journal of Computer Science and Engineering / SSRN preprint*.
7. Student placement prediction using various machine learning models (IJFMR / IJFMR PDF).