STUDENT PLACEMENT PREDICTION USING MACHINE LEARNING

Shreyas Harinath [1], Aksha Prasad [2], Suma H S[3], Suraksha A[4], Tojo Mathew [5]

[1,2,3,4]BE, Department of CSE, NIE Mysore, Karnataka, India
[5]Assistant Professor, Department of CSE, NIE Mysore, Karnataka, India

---***---

Abstract – *Placement of scholars is one in every of the vital activities in academic establishments. Admission and name of establishments primarily depends on placements. Hence all institutions strive to Strengthen placement department. The main Objective of this paper is to analyze previous year's student's historical data and predict placement possibilities of current students and aids to increase the placement percentage of the institutions. This paper presents a recommendation system that predicts whether the current student will be placed or not, if the student is placed the company is also predicted based on the data of previously placed students. Here we use two different machine learning classification algorithms, namely Naive Bayes Classifier and K-Nearest Neighbors [KNN] algorithm. These algorithms independently predict the results and we then compare the efficiency of the algorithms, which is based on the dataset. This model helps the position cell at intervals a corporation to spot the potential students and concentrate to and improve their technical and social skills.*

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

1. INTRODUCTION

Nowadays educational institutes are growing in high numbers. Aim of every higher educational institute is to get their students a well-paid job through their placement cell. One of the largest challenges that higher learning establishments face nowadays is to boost the placement performance of scholars. The placement prediction is additional complicated once the quality of instructional entities increase. One of the effective ways to address the challenges for improving the quality is to provide new knowledge related to the educational processes and entities to the managerial system. With the machine learning techniques the information are often extracted from operational and historical knowledge that resides at intervals the academic organization's databases exploitation. The information set for system implementation contains data regarding past data of scholars. These knowledges square measure used for coaching the model for rule identification and for testing the model for classification. The prediction of placement status that students are most likely to achieve will help students to put in more hard work to make appropriate progress in stepping into a career in various technical fields. It will also help the teachers as well as placement cell in an institution to provide proper care towards the improvement of students in the duration of

course. A high placement rate is a key entity in building the reputation of an educational institution. Hence such a system has a significant place in the educational system of any higher learning institution. We use Naive Bayes and K-Nearest neighbors [KNN] machine learning module to provide efficient and accurate results.

1.1 Prediction System

In this paper the focus on machine learning technique to predict placement status of the student provided through text input. The placement prediction is done by machine learning using Naïve Bayes and K-nearest neighbor (KNN) algorithm. The algorithm considers the parameters such as USN, Tenth and PUC/Diploma results, CGPA, Technical and Aptitude Skills.

1.2 Naive Bayes Classifier

The Naive Bayes Classifier is very effective on many real data applications. The performance of Naïve Bayes usually benefits from an precise estimation of univariate conditional probabilities and from variable selection.

1.3 K-nearest neighbor (KNN)

KNN is a simple algorithm which uses entire dataset during its training phase whenever prediction is required for unseen data. It searches through entire training dataset for k- most similar instances and data with most similar instance are returned.

1.2.1 Features of KNN

- KNN stores the complete training dataset which it uses as its representation.
- It makes predictions just-in-time by calculating the similarity between an input sample and each training instance.
- It works on similarity measures.

2. RELATED WORK

Senthil Kumar Thangavel, Divya Bharathi P and Abhijith Shankar [1] conducted a study to predict student placement status using two attributes, areas and CGPA results. They made use of Decision Tree Learning, SCI-Kit leaning in machine Learning here they use only two parameters such as CGPA and arrears used algorithm takes more time for prediction not efficient.

Wilton W.T. FOK, Y.S. He, H.H Au Yeung and K.Y. Law [2] Conducted a study to predict suitable course for the students, based on their behavior using Neural Network Technique. TensorFlow engine includes number of intermediate node and number of deep learning layers are adjusted and compared.

Machine Learning deals with the development, analysis and study of algorithms that can automatically detect patterns from data and use it to predict future data or perform decision making [3]. Machine learning does its functionality by creating models out of it [4]. Machine Learning has become widespread and has its applications in the field of bioinformatics, computer vision, robot locomotion, computational finance, search engine etc.

3. METHODOLOGY

Naive Bayes

This is an easy technique for building classifiers: models that assign class labels to downside instances, painted as vectors of feature values, where class labels are drawn from a few finite set. There is no single algorithm for training such classifiers, however a family of algorithms which is based on a standard principle: all Naive Bayes classifiers assume that value of a particular feature is independent of the value of other feature, given class variable. For example, a fruit could also be thought of be an apple if it is red, round, and about 11 cm in diameter. A Naive Bayes classifier considers every of these options to contribute severally to the likelihood that this fruit is AN apple, no matter any potential correlations between the colour, roundness, and diameter features.

Working of Naive Bayes Algorithm

Step 1: Scan the dataset (storage servers) retrieval of required data for mining from the servers such as database, cloud, excel sheet etc.

Step 2: Calculate the probability of every attribute value. [n, n_c, m, p] Here for each attribute we calculate the probability of occurrence using the following formula. (mentioned in the next step). For each class (Course) we should apply the formulae.

Step3:P(attributevalue(ai)/subjectvaluevj)=(n_c+mp)/(n+m) apply the above formulae
Where:
n = no. of training examples for which v = vj
nc = no. of examples where v = vj and a = ai
p = a priori estimate for P(aivj)
m = the equivalent sample size

Step 4: Multiply the probabilities by p for each class, here we multiple the results of each attribute with p and final results are used for classification.

Step 5: Compare the values and classify the attribute values to 1 of the predefined set of class.

K-nearest Neighbors classifier (KNN)

K-Nearest Neighbors (KNN) [3] is a simple, lazy and nonparametric classifier. KNN is favoured when all the features are continuous. KNN is additionally referred to as case-based reasoning and has been utilized in several applications like pattern recognition, statistical estimation. Classification is obtained by identifying the nearest neighbors to determine the class of an unknown sample. KNN is picked over other classification algorithms because of its high convergence speed and ease. Figure below shows nearest neighbors classification. KNN classification has two stages.
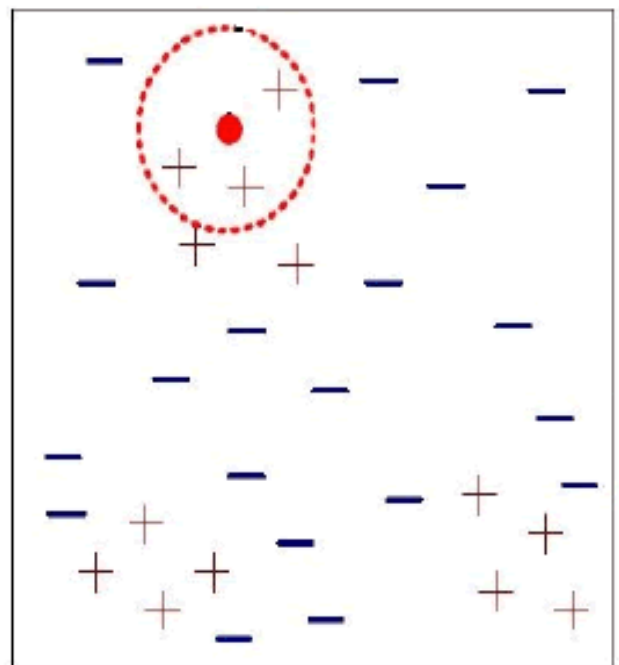
1) Find the k number of instances within the dataset that's closest to instance S

2) These k number of instances then vote to determine the class of instance S

The Accuracy of KNN is dependent on the distance metric and K value. Various ways of measuring the distance between 2 instances are cosine, Euclidian distance.

Working of KNN Algorithm

The suggested method here aims to enhance the performance of KNN classifier for disease prediction. Algorithm for our proposed method is shown below

Step 1: Input the data set.

Step 2: Apply KNN algorithm for input parameters.

Step 3: Euclidean distance between existing parameters and newly entered is calculated.

Step 4: Based on similarity measures output is predicted.

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 $X_i = x_1, x_2, x_3, \ldots\ldots, x_n$

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

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

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

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

4. FUTURE ENHANCEMENT

The future enhancements of the project is to focus on to add some more parameters to predict more efficient placement status. We can also enhance the project by predicting some solutions or suggestions for the output generated by system.

5. CONCLUSION

Student Placement Predictor is a system which predicts student placement status using machine learning techniques. Many research papers are there related to educational sector, all these papers mainly concentrate on student performance predictions. All these predictions help the institute to improvise the student performance and can come up with 100% results. Many of the previous research papers concentrate on a less number of parameters such as CGPA and Arrears for placement status prediction which leads to les accurate results, but proposed work contains many educational parameters to predict placement status which will be more accurate.

ACKNOWLEDGEMENT

REFERENCES

[1] "Student Placement Analyzer: A Recommendation System Using Machine Learning" 2017 International Conference on advanced computing and communication systems (ICACCS-2017), Jan 06-07,2017, Coimbatore, INDIA.

[2] "Prediction Model for Students Future Development by Deep Learning and TensorFlow Artificial Intelligence Engine" 2018 4th IEEE International Conference on Information Management.

[3] Kohavi, R and F, Provost (1998). Machine Learning 30:271-274.