Ravi Kumar V ,INDIA / International Journal of Research and Computational Technology
Vol.14 Issue.1 *Free Journal Publication* *Pages: 11- 18*
ISSN: 0975-5662, *June, 2022* *www.ijrct.com*

STUDENT PLACEMENT PREDICTION USING ML

RAVI KUMAR.V[1], Email:ravivannoj@tkrcet.com
#**KALUVALA SOUMYASREF[2]**,Email:kaluvalasoumyasree@gmail.com
P. MANIRADHAR[3], Email:radharash72@gmail.com
#**SULTAN SAIKUMAR[4]**,Email:sultansaikumar61@gmail.com
K.SAIKIRAN[5], Asst. Professor, TKR College of Engineering and Technology, Hyderabad,INDIA.
#BTech Student, TKR College of Engineering and Technology, Hyderabad,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. 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 Random Forest Classifier, Gaussian Naive Bayes Classifier, Logistic Regression algorithm to 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.

Keywords: recommendation system,r Random Forest Classifier, Gaussian Naive Bayes Classifier, Logistic Regression

1.INTRODUCTION

According to statistics 1.6 million students pass from CS department every year. The demand for skilled and qualified students is rising day by day. Thus the company use a good amount of capital in recruiting students from in-campus and off-campus because number of skilled and qualified students are very low. Colleges and Institution needs to focus on practical knowledge of real world rather than completing their syllabus. Placements are the biggest opportunities in the life of a student and they need to be fully prepared while attempting it. Placement Predictor system helps in predicting whether a student will get placement or not. This system can also be helpful for identifying the areas where student needs to work on for placement. This system uses student's details like academic marks, coding skills, etc. This system uses previous year placement statistics and student dataset for the placement prediction so the placement cell of the organization could set up a placement anticipated rundown for the present students. Along these lines it is

important to direct an investigation on different placement prediction frameworks. This paper shows an overview on various placement prediction framework models and its application for the students.

2. Related works

Generally all the academic established institutions are having placement department. Some of the institutions are maintain a common database in general applications like excel sheets and databases uses their own software applications. So that in existing systems they does not have a proper analysis software applications. Some third parties conducted a study using Decision Tree Learning, SCI-kit learning in machine learning using two attributes, areas and CGPA results takes more time not efficient. And other conducted to predict course 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.

Disadvantages:

1. They used normal conditional statements of getting information

2. They don't have proper algorithms for analyzing data for new placements.

3. Methodology

We proposed 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 KNearest Neighbors [KNN] algorithm. Machine learning does its functionality by creating models out of it by predicting the future data.

Advantages:

1. 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.

4. IMPLEMENTATION

- Gathering Data
- Data preparation
- Data Wrangling

- Analyze Data
- Train the model
- Test the model
- Deployment

a. Gathering Data:

Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.

In this step, we need to identify the different data sources, as data can be collected from various sources such as files, database, internet, or mobile devices. It is one of the most important steps of the life cycle. The quantity and quality of the collected data will determine the efficiency of the output. The more will be the data, the more accurate will be the prediction.

This step includes the below tasks:

- Identify various data sources
- Collect data
- Integrate the data obtained from different sources

By performing the above task, we get a coherent set of data, also called as a dataset. It will be used in further steps.

b. Data preparation

After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.

In this step, first, we put all data together, and then randomize the ordering of data.

This step can be further divided into two processes:

- Data exploration:

It is used to understand the nature of data that we have to work with. We need to understand the characteristics, format, and quality of data.

A better understanding of data leads to an effective outcome. In this, we find Correlations, general trends, and outliers.

- Data pre-processing:

Now the next step is preprocessing of data for its analysis.

c. Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to

make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.

It is not necessary that data we have collected is always of our use as some of the data may not be useful. In real-world applications, collected data may have various issues, including:

- Missing Values
- Duplicate data
- Invalid data
- Noise

So, we use various filtering techniques to clean the data.

It is mandatory to detect and remove the above issues because it can negatively affect the quality of the outcome.

d. Data Analysis

Now the cleaned and prepared data is passed on to the analysis step. This step involves:

- Selection of analytical techniques
- Building models
- Review the result

The aim of this step is to build a machine learning model to analyze the data using various analytical techniques and review the outcome. It starts with the determination of the type of the problems, where we select the machine learning techniques such as Classification, Regression, Cluster analysis, Association, etc. then build the model using prepared data, and evaluate the model.

Hence, in this step, we take the data and use machine learning algorithms to build the model.

e. Train Model

Now the next step is to train the model, in this step we train our model to improve its performance for better outcome of the problem.

We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.

f. Test Model

Once our machine learning model has been trained on a given dataset, then we test the model. In this step, we check for the accuracy of our model by providing a test data set to it.Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

g. Deployment

The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system.

If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not. The deployment phase is similar to making the final report for a project.

Algorithms and Techniques

Three supervised learning approaches are selected for this problem. Care is taken that all these approaches are fundamentally different from each other, so that we can cover as wide an umbrella as possible in term of possible approaches. For example- We will not select Random Forest and Ada Boost together as they come from the same family of 'ensemble' approaches:

For each algorithm, we will try out different values of a few hyperparameters to arrive at the best possible classifier. This will be carried out with the help of grid search cross validation technique. The algorithms are described below:

Random Forest Classifier:

on_estimators(number of trees in a forest) o max_depth(maximum depth of one single tree) o max_features(decides how many features are to be used) o oob_score(decides whether to include out-of-bag or prediction error)

Gaussian Naive Bayes Classifier

Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. The technique is easiest to understand when described using binary or categorical input values. The representation for naive Bayes is probabilities. A list of probabilities are stored to file for a learned naive Bayes model. This includes:

Class Probabilities: The probabilities of each class in the training dataset. Conditional Probabilities: The conditional probabilities of each input value given each class value.

. **Logistic Regression:**

Since the outcome is binary and we have a reasonable number of examples at our disposal compared to number of features, this approach seems suitable. At the core of this method is a logistic or sigmoid function that quantifies the difference between each

prediction and its corresponding true value. When presented with a number of inputs, it assigns different weights to features (based on their relative importance).

Since for this data it already knows the output beforehand, it continuously adjusts the weights such that when these weights summed up with their features are introduced in the logistic function, the results are as near as possible to the actual ones. Once presented with a test value, it again inserts the value into our logistic function and returns the output as a number between 0 and 1, which represents the probability of that test value being in a particular class.

Findings and Suggestions

The dataset for this problem is the ILPD (Indian Liver Patient Dataset) taken from the UCI Machine Learning Repository. Region specified in this dataset is of Andhra Pradesh of year 2017.

As our dataset is small it's training dataset is similar to test dataset so we cannot rely on this model for predicting accuracy for large dataset. We need more precised data set containing larger values and attributes to classify values and get best accuracy

Conclusion:

Predicting the placement of a student gives an idea to the Placement Office as well as the student on where they stand. Not all companies look for similar talents. If the strengths and weaknesses of the students are identified it would benefit the student in getting placed. The placement Office can work on identifying the weaknesses of the students and take measures of improvement so that the students can overcome the weakness and perform to the best of their abilities. Thus the key lies in assessing the capabilities of the student in the right areas and subjecting them to the right training.

Future scope:

The above work was carried out with respect to the Information Science Engineering and Computer Science Engineering branch of our college. Further work can be carried out by applying other algorithms that could lead to improvement in results, also different Skill sets for above mentioned branches and on data of different streams of engineering. The key to this would be to identify the aspects of these engineering branches that would test the student's knowledge in getting placed in a core company for their respective branches.

References:

[1] Manoj K Shukla, PranayRambade, Jay Torasakar, RakeshPrabhu, Prof. DeepaliMaste, "Students Placement Prediction Model Using Logistic Regression", International Journal Of Engineering Research & Technology (IJERT), 2017.

[2] Ravi Kumar Rathore and J. Jayanthi, "Student Prediction System For Placement Training Using Fuzzy Inference System", ICTACT Journal On Soft Computing,2017

[3] MangasuliSheetal, Prof. SavitaBakare, "Prediction of Campus Placement Using Data Mining Algorithm-Fuzzy logic and K nearest neighbor", International Journal of Advanced Research in Computer and Communication Engineering, 2016.

[4] Bharambe, Yogesh, "Assessing employability of students using data mining techniques", Advances in Computing, Communications and Informatics (ICACCI), 2017.

[5] P. Khongchai, P. Songmuang, "Random Forest for Salary Prediction System to Improve Students Motivation", Signal-Image Technology & Internet-Based Systems (SITIS), 2016.

[6] Hitarthi Bhatt, Shraddha Mehta, Lynette R. D'mello, "Use of ID3 Decision Tree Algorithm for Placement Prediction", (IJCSIT) International Journal of Computer Science and Information Technologies, 2015.

[7] Dammalapati Rama Krishna, Bode Prasad, TekiSatyanarayana Murthy, "Placement Prediction Analysis in University Using Improved Decision Tree Based Algorithm", International Journal of Computer Trends and Technology (IJCTT), 2014.

[8] SyedJunaidRahmathulla, "A Combined Clustering and Classification Approach for Predicting Placement Chance", International Journal of Engineering Research & Technology (IJERT), 2016.

[9] Ramanathan.L, Swamalatha P, D. GaneshGopal, "Mining educational data for students' placement prediction using sum of difference method", International journal of computer application, 2014.

[10] PPS - Placement Prediction System using Logistic Regression ,Ajay Shiv Sharma, Swaraj Prince, ShubhamKapoor, Keshav Kumar,2014 .

[11] Osmanbegoviü, Edin, and MirzaSuljiü. "Data mining approach for predicting student performance."Economic Review 10, no. 1 (2012).

[12] Kabakchieva, Dorina. "Predicting student performance by using data mining

methods for classification." Cybernetics and Information Technologies 13, no. 1 (2013): 61-72.