INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1253)
Available online at: https://www.ijariit.com

Smart system for student placement prediction

Kavya G.
kavyag.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Rajanukunte, Karnataka

Pranitha Y.
pranithay.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Rajanukunte, Karnataka

Sanjana A.
sanjanaarunkumar.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Rajanukunte, Karnataka

Sirisha D. G.
sirishadg.17cs@saividya.ac.in
Sai Vidya Institute of Technology, Rajanukunte, Karnataka

Mamatha A.
mamatha.a@saividya.ac.in
Sai Vidya Institute of Technology, Rajanukunte, Karnataka

ABSTRACT

Its goal is to develop college-level forecasting automation, predict the chances of students being placed in information technology companies, and help screen candidates before the hiring process begins. lt involves using machine learning algorithms as a basic model to classify students into appropriate groups, so the results will help them improve their skills and other ways of thinking. The same results are also compared with the results obtained from other models such as logistic regression, random forest and SVM to obtain the optimal and so on. Placements are considered to be very important for each and every college. The main objective of this model is to predict whether the student gets placed or not in campus recruitment. A high placement rate is a key entity for any educational institution. Hence such a system has a significant place in the educational system of any learning institution.

Keywords: *Classification, Machine Learning, Automation, Logistic Regression, Student Placement.*

1. INTRODUCTION

Predicting student placement based on student characteristics is a tedious educational process. The existing placement prediction system only considers academic performance to predict whether a student will be placed. Judging undergraduates based on academic performance alone is unfair to students, because undergraduates may have good talents, skills, and communication skills, but unfortunately may not be good at academics.

The proposed system determines the chances of a student getting placed in a company by using classification algorithms such as K-nearest neighbour, logistic regression, SVM and Decision Tree. Proposed system is an automation for placement prediction. The Proposed methodology analyses the previous year's student's data and determines placement chances of "current students" and placement percentage of the institution. The Proposed system mainly concentrates on student knowledge, skills and attitude. The system helps in improving the placement rate of an institution thereby can act as a key element in improving the reputation of the institution.

2. LITERATURE SURVEY

- **Shreyas Harinath, Aksha Prasad, Suma H S, Suraksha A, Tojo Mathew:** Student Placement Predictor is a system which predicts student placement status using machine learning techniques. Here they use two different machine learning classification algorithms, namely Naive Bayes Classifier and K-Nearest Neighbour [KNN] algorithm. These algorithms independently predict the results and then compare the efficiency of the algorithms.
- **Abhishek S. Rao, Aruna Kumar S V, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B, Prashanth Gouda:** Here they proposed a system to predict student's placement with the help of Data Mining approach. In this work, three best-suited classification algorithms like ANN, KNN and SVM were utilized. From the experimental analysis, it was evident that the results obtained for the ANN algorithm were best compared to the other two classifiers when 80% of the data was trained and 20% of the data was tested using the Tanh activation function.
- **Alfiani, Ardita Permata, and Febriana Ayu Wulandari:** In this paper, the author emphasized on the mapping of students' performance patterns using a data mining approach with the help of a case study in Indonesia. K-means clustering was used for revealing the hidden patterns from the data. A total of 306 valid cases were used for the study. Various attributes selected were GPA, Lecture grades, and Lab grade. The student's performance was classified into three categories i.e. low, average and smart .
- **Liu, Yang, et al:** This paper highlights a database application for huge data management in colleges. Decision Tree and association rules algorithms were used for data pre-processing, in which student's scores were analyzed using decision tree

algorithm and consumption, psychological status was analyzed using association rule algorithms. The performance was classified as poor, and excellent. Finally, the designed system could automatically show the results using the data mining algorithms chosen for the study.

- **Sheetal, M. B, Savita, Bakare :** This paper highlights the importance of campus placement in a student's life and how the prediction of the campus placement can be achieved with the help of data mining algorithms like fuzzy logic and K-Nearest Neighbors (KNN). Results indicated that the accuracy and execution time of KNN was 97.33% and 13458 (msec) when compared to fuzzy logic with accuracy 92.67% and execution time of 450 (msec).
- **Sumitha, R., E. S. Vinothkumar, and P. Scholar:** In this paper, the author designed a data model for better management of student's performance, using a senior student's dataset. The J48 algorithm was used and the accuracy of the model was compared with other data mining algorithms. High accuracy of 97.27% was achieved by the J48 algorithm when compared to Naïve Bayes and multilayer perceptron i.e. 85.92% and 94.94% respectively. Selected features for the study were CGPA, arrears, attendance, PUC marks, Engineering cut off, Medium of education and type of board. Performance classified notations given were good average and best.

3. PROPOSED SYSTEM AND METHODOLOGY

The design consists of various strategies used in the prediction process; the details of the system model is present in Figure 1.

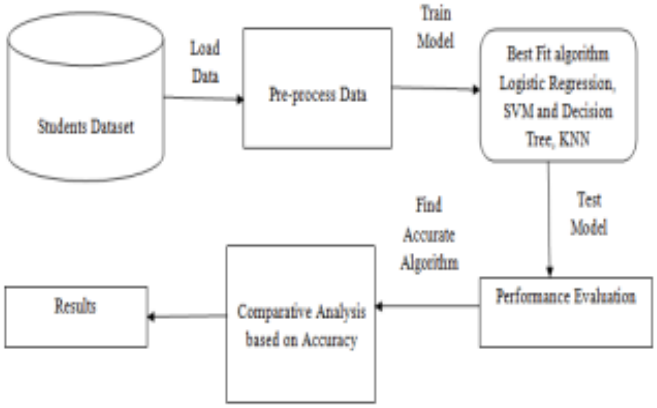

Fig 1: Flow Diagram of Student Placement prediction

Training and Testing of Data:
Data set is split into Train Data and Test Data in the ratio 80:20.

Data set: This model has used Placement data sets from online websites like Kaggle. A placement dataset from Kaggle is used to train the model and then another Twitter data set is used for getting the result. The Dataset used in the project has 600 rows of data. The attributes taken into consideration are SSC_Percentage, HSC_Percentage, UG_Percentage, Logical Reasoning Score, Aptitude Score, Programming Score, Communication skills Score.

Data Pre-processing: When the data is considered, always a very large data set with a large no. of rows and columns will be noted. But it is not always the case that the data could be in many forms such as Images, Audio and Video files, Structured tables, etc. The machine doesn't understand images or video, text data as it is, Machine only understands 1s and 0s.

Data Collection: The data set includes the academic and even technical and cognitive skill.The dataset consists of 600 individual datasets that are considered from the previous year students.The attributes taken into consideration are SSC_Percentage, HSC_Percentage, UG_Percentage, Logical Reasoning Score, Aptitude Score, Programming Score, Communication skills Score.

Data Preparation: Format and architect the information and data into ideal arrangement, excluding the important and significant highlights and performing reduction of dimensionality.

Training(fitting stage): This is the place the Machine Learning algorithms perform calculation and really learns by indicating the information and data that has been gathered precisely and arranged.

Software Specification: Operating system used is Windows XP / 7, Coding Language is Python, HTML and Version of python used is Python 3.6.8, IDE is Python 3.6.8 IDLE, ML Packages are Numpy, Pandas,Sklearn, Matplotlib, Flask, PymySql, NLTK. ML Algorithms are Logistic Regression, Random forest, KNN, Decision tree, SVM and Naïve Bayes and Other Requirements required are Notepad and XAMPP Control Panel.

4. ALGORITHMS

1. Logistic Regression: Logistic regression is another method borrowed from the field of machine learning statistics. which is the way to go for binary classification problems. Logistic regression is named after the logistic function, which is the function used at the core of the method, also called the sigmoid function, which was developed to explain the attributes of population growth in ecology and rapidly. Increases to maximize transport capacity. Environmental. An S-shaped curve that can take the actual value and map it to a value between 0 and 1.

2. Naive Bayes : The naive Bayes classifier is a generative model for classification. Before the advent of deep learning and it's easy-to-use libraries, the Naive Bayes classifier was one of the widely deployed classifiers for machine learning applications. Despite its simplicity, the naive Bayes classifier performs quite well in many applications.

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

With Bayes Theorem one can find probability that happens because B has occurred. Here is B-proof and A is the hypothesis. The assumptions made here are predictions / characteristics are independent. It is the presence of a particular feature that does not affect the other side. Therefore, it is called innocent.

3. Random Forest Classifier: Random Forest is also called a supervised learning algorithm. Both classification and regression are available. It is also the most flexible and easy to use algorithm. A forest is made up of trees. A random forest builds a decision tree
Here is an example of the logistic regression equation:
y = e^(b0 + b1*x) / (1 + e^(b0 + b1*x))on a randomly selected sample of data, makes predictions from each tree, and selects the optimal solution through voting.It technically is a method (based on the divide-and-conquer approach) of decision trees that is generated on a randomly data split dataset.

4. K-Nearest Neighbour (KNN): K-nearest neighbour is the non-parametric lazy algorithm. The nearest neighbours are

selected based on Euclidean distance calculated between x and y vectors given in the equation (2). The result of KNN varies for different values of K . A large value of K will cause overlapping in classes, while a smaller value of K increases computations.

$$\text{Euclidean Distance} = \sqrt{\sum_{i=1}^{k} (x_i - y_i)^2}$$

5. Decision Tree Classifier: The tree structure is similar to the flowchart in which the internal nodes of the decision tree indicate the function (or attribute), the branch indicates the decision rule, and each leaf node shows the result.

The top node of the decision tree is called the root node. Learn how to split based on attribute values. Call recursive partitioning to recursively split the tree. This flowchart-like structure is useful for decision making. Therefore, decision trees are very easy to understand and also to interpret. The basic idea behind any decision tree algorithm is as follows:

a. Select the best decision attribute using ASM to split the record.
b. Make the attribute as a decision tree node and also break the dataset into a smaller set of subsets.
c. Now start the tree building process by repeating the steps recursively for each child until one of the conditions matches :
   i. all the tuples are now belonging to the same value of attributes.
   ii. There seems to be no more remaining attributes.
   iii. There seems to be no more instances.

6. Support Vector Machine(SVM): Support vector or SVM machine is one of the most commonly monitored learning algorithms used to classify and regression issues. Mostly, however, it is used to classify problems in learning machines. The aim of the SVM algorithm is to prepare the best line or decision-making limit that can separate the N-dimensional space in classes, so it can easily place new data points in the future category. This best decision-making limit is called a hyperplane. SVM Select Extreme Points / Vectors to create Super Flat. These extreme cases are referred to as support vectors, and because the algorithm is referred to as a carrier vector.

The SVM classifier offers better accuracy and faster predictions. It also uses less memory because it uses a subset of educational points during the decision-making phase. SVMs work clearly with separation margins and higher dimensional spaces.

5. RESULTS

The model has been trained using multiple algorithms to check and compare the results for greater accuracy. After knowing the accuracy of all the algorithms, select the one with highest accuracy as the finalized algorithm.

Table-1: Accuracy Table

SL.No	Models	Accuracy
1	Gaussian Naive Bayes	93.33%
2	Logistic regression	95.00%
3	Random Forest	94.17%
4	KNN	87.50%
5	Decision Tree	91.67%
6	SVC	79.17%

```
Accurcay of LogisticRegression is 95.00%
Accurcay of RandomForestClassifier is 94.17%
Accurcay of KNeighborsClassifier is 87.50%
Accurcay of SVC is 79.17%
Accurcay of GaussianNB is 93.33%
Accurcay of DecisionTreeClassifier is 91.67%
>>>
```

Fig 2: Accuracy of algorithms used

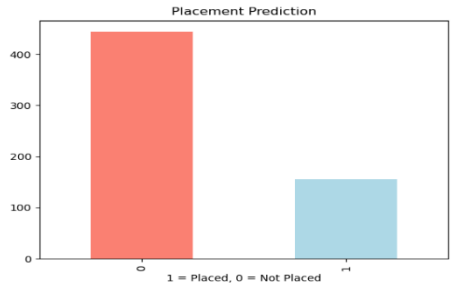

Fig 3: Graphical Representation Of Placement Prediction

In the Front end, the user will enter the placement data. The data is taken as input to the finalized algorithm to predict the output. The Output will be either the student will be placed or not placed.

6. CONCLUSION

Student placement prediction is a system that uses machine learning technology to predict placement status. Much of the research work is related to the education sector; all these articles will focus primarily on the prediction of student performance. All these predictions help the college improvise in student performance and enable them to achieve good results. Many previous research papers focused on a relatively small number of parameters, such as CGPA and Lags used to predict the position state, resulting in inaccurate results, but the proposed work contains many educational parameters for predicting the position state. precise.

So a system is proposed with the help of machine learning technology and algorithms, such as logistic regression, KNN, SVC, random forest, decision tree and naive Bayes, to predict whether students will be positioned based on different parameters, such as As an academic achievement, Communication skills, technical skills, etc. The proposed system predicts whether students will be placed or not with a good 95% accuracy given by logistic regression.

7. ACKNOWLEDGEMENT

We are grateful for our Guide and Assistant Professor Mamatha A's support, advice and encouragement.

8. REFERENCES

[1]. Shreyas Harinath ,Aksha Prasad,Suma H S, Suraksha A,Tojo Mathew "STUDENT PLACEMENT PREDICTION USING MACHINE LEARNING "Volume: 06 Issue: 04 | Apr 2019 .

[2.] Abhishek S. Rao, Aruna Kumar S V, Pranav Jogi, Chinthan Bhat K, Kuladeep Kumar B, Prashanth Gouda "Student Placement Prediction Model: A Data Mining Perspective for Outcome-Based Education System" International Journal of

Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019.

[3.] Alfiani, Ardita Permata, and Febriana Ayu Wulandari. "Mapping a student's performance based on a data mining approach (a case study)." Agriculture and Agricultural Science Procedia 3 (2015): 173-177.

[4.] Liu, Yang, et al. "The Application of Data Mining Techniques in College Students Information System." 2018 International Conference on Computer Science, Electronics and Communication Engineering (CSECE 2018). Atlantis Press, 2018.

[5]. Sheetal, M. B, Savita, Bakare. "Prediction of Campus Placement with Data Mining Algorithm-Fuzzy logic and K-nearest neighbor." International Journal of Advanced Research in Computer and Communication Engineering 5.6 2016: 309-312.

[6.] Sumitha, R., E. S. Vinothkumar, and P. Scholar. "Prediction of Students Outcome Using Data Mining Techniques." Int. J. Sci. Eng. Appl. Sci 2.6 (2016): 132-139

BIOGRAPHY

Mamatha A.
Assistant Professor, Dept of CSE,
Sai Vidya Institute of Technology, Rajanukunte, Bangalore, Karnataka, India

Kavya G.
Student,
Sai Vidya Institute of Technology, Rajanukunte, Bangalore, Karnataka, India

Pranitha Y
Student,
Sai Vidya Institute of Technology, Rajanukunte, Bangalore, Karnataka, India

Sanjana Arunkumar
Student,
Sai Vidya Institute of Technology, Rajanukunte, Bangalore, Karnataka, India

Sirisha D. G.
Student,
Sai Vidya Institute of Technology, Rajanukunte, Bangalore, Karnataka, India

