“Placement Prediction and Analysis using Machine Learning”

Rushikesh Joshi
Computer Engineering
VPSCET Lonavala
Student,SPPU India
naymish2002@gmail.com

Vijay Shilwant
Computer Engineering
VPSCET Lonavala
Student , SPPU India
shilwantvijay123@gmail.com

Roshan Dandge
Computer Engineering
VPSCET Lonavala
Student ,SPPU India
roshandandge25@gmail.com

Shreyash Rajput
Computer Engineering
VPSCET Lonavala
Student,SPPU India
shreyashdr@gmail.com

Dr. Manav A Thakur
Computer Engineering
VPSCET Lonavala
Principal ,SPPU India
vpscet@gmail.com

ABSTRACT:

We have designed a campus placement prediction system aimed at assessing a student's likelihood of securing employment through campus recruitment. This predictive model incorporates multiple parameters to gauge a student's skill level. Some of these parameters are drawn from the college's records, including academic performance, CGPA, attendance, and more, while others are derived from assessments conducted within the placement management system. By amalgamating these data points, our model can provide accurate predictions regarding a student's potential placement in a company.

Furthermore, we leverage data from previous group of students to train our model, employing educational data mining techniques to access authentic historical data from our college's alumni. This approach enhances the efficacy of our machine learning model when making predictions specific to our institution.

Keywords— Machine Learning, Classification, Result based instruction, Placement Prediction Introduction.

INTRODUCTION

Placements are crucial for colleges as they reflect the institution's success. Our approach focuses on predicting andanalyzing these essential campus placements to enhance students' chances. The primary goal is to forecast whether students will secure campus recruitment. To achieve this, weconsider students' academic performance in key subjects such as Data Structures and Algorithms, Object-Oriented Programming, Database Management System, Operating System, Computer Network Security, and Cloud Computing, which are pivotal for campus placements. This approach encourages students to excel academically and develop additional skills for better placement opportunities. Additionally, it helps institutions improve their placement records, attract new admissions, and enhance their reputation.

Our proposed system maintains student details and generates lists of candidates for companies seeking recruits with specific skills. The project's objective is to minimize errors associated with manual processes, save time, and empower students to identify their strengths and weaknesses to optimize their placement prospects.

Machine Learning :

Machine learning, on the other hand, is an automated process that enables machines to solve problems with little orno human input, and take actions based on past observations. While artificial intelligence and machine learning are often used interchangeably, they are two different concepts.

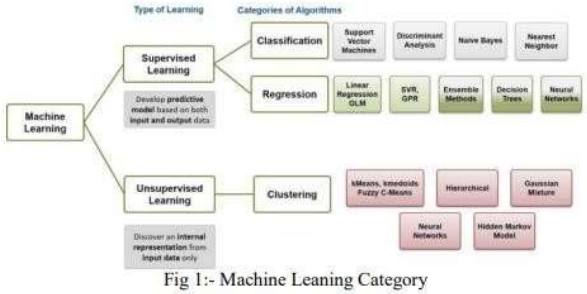

Fig 1:- Machine Leaning Category

From the previously mentioned AI models Supervised learning is utilized in this undertaking.

2. LITERATURE SURVEY

Sharma et al. Al. We have developed a placement prediction system. H. PPS, using a logistic regression model. He considered characteristics such as entry grade, secondary level II performance, and performance in subjects in different semesters. The dataset used here is from his GuruNanak Dev Engineering College (GNDEC), Ludhiana. The accuracy of this model was approximately 83.33%. [2]

Elayidom et al. Al. We built a multi-way decision tree using various parameters such as industry, sector, gender, and rank. The datasets used are received from the National Technical Manpower Information System (NTMIS) via the Nodal Center. The accuracy of this model was 80%. [5]

Nagaria et al. Al. He used his random forest model and took into account various parameters such as degree type, work experience, exam percentage, specialization, and MBA percentage. The dataset used is from his Kaggle. This model had the highest accuracy of 85%. [6]

S. Venkachalam et al. Al. We designed a fuzzy inference system to predict campus placement using a naive Bayes algorithm. Data sets are created using primary and secondary data collection sources. This model had the highest accuracy of 86.15%.[7]

Manvitha et al. Al. The random forest model was designed considering various parameters such as credit score, delinquency, payment status, b.tech %, etc. The dataset is collected by the placement department of Sreenidhi University of Science and Technology. This model had the highest accuracy of 86%.[8]

OVERVIEW OF THE SYSTEM

Under the current system, India trains 1.5 million engineers every year and the need for experts in the IT field is increasing. However, many students are unaware of industryneeds, leading to a low number of graduates meeting business requirements. The student's position poses a significant challenge. Educational institutions should maximize internship opportunities, and recruiting departments and teachers should prepare students to meet specific company requirements. The job placement prediction system can evaluate a student's suitability for a specific job. In the proposed system, IT companies invest huge amounts of money in recruiting students. To reduce these costs, effectivestudent filtering can be achieved using deep learning tools. Deep learning involves identifying patterns in data through various analytical methods. Educational institutions can leverage this data mining capability to understand the company's recruitment history and student data, thereby allowing the attribution cell to predict current student placement. This article studies different models of location prediction systems and their applications for students.

II. METHODOLOGY

The steps involved in this system are as follows,

A. Data Acquisition:

The campus placement dataset is collected from Kaggle website. Here is the link for the dataset: https://www.kaggle.com/benroshan/factors-affecting-campus placement?select=Placement_Data_Full_Class.csv
The dataset consists of various attributes such as serial number, Gender, SSC Percentage, SSC Table . - Central/Other, HSC Percentage, HSC Specialization, Diploma Percentage, Diploma Specialization, Work Experience, E-Test Percentage and Salary Expectation.

B. Handling categorical data:

Since we cannot deal with categorical values directly ,mapping is done for attribute. Handling categorical data in machine learning is a crucial step, as most machine learning algorithms work with numerical data. Categorical data represents discrete, non-numeric values such as names, labels, or categories.

C. Pre-processing the data:

Data preprocessing is a critical step in machine learning, as the quality of your data and how it's prepared can have a significant impact on the performance of your models. Data preprocessing involves various tasks aimed at cleaning, transforming, and organizing data to make it suitable for training machine learning models.

D. Handling missing values:

In our data set, missing values appear only in the salary column because these values correspond to students who were not placed in any intake. So, assume that the missing values in the Salary column

are 0 and replace them with 0 using the fillna(0,inplace=True) function in Python.

E. Manage classification data:

Since we cannot deal with categorical values directly, Mapping is done for attributes with categorical values. The Gender attribute has the values M (Male) and (M). This M is replaced by 0 and F is replaced by 1. The SSC and HSC tag attributes have the values “Central” and “Other”. Here Central is replaced with 1 and Other is replaced with 0. The Work Experience attribute has the values "Yes" and "No". Here, “Yes” is replaced by 1 and “No” is replaced by 0.

The Status attribute has the values “Set” and “Unset”. Here, “Set” is replaced by 1 and “Not Set” is replaced by 0. This is achieved using the map function in Python.ibutes having categorical values.

Status attribute has values ‘Placed’ and ‘Not Placed’. Here, ‘Placed’ is replaced by 1 and ‘Not Placed’ is replaced by 0. This is achieved through map function in Python.

For e.g.,

```
x df['gender']=df['gender'].map({'M':0,'F':1})x
df['ssc_b']=df['ssc_b'].map({'Central':1,'Others':0})x
df['workex']=df['workex'].map({'Yes':1,'No':0})
```

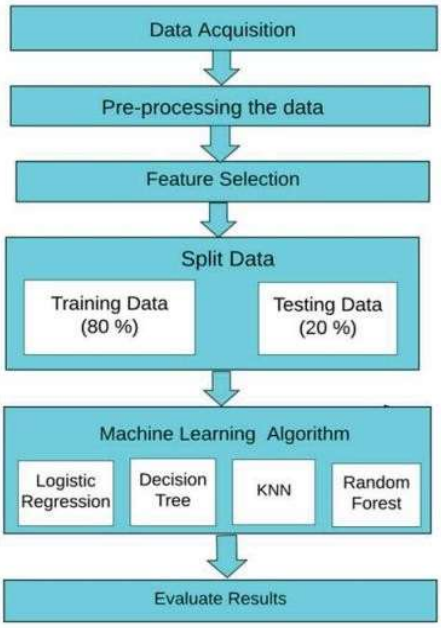

Fig.1 .Architecture Diagram

F. Feature Selection:

Here, various features are visualized to understand their correlation with the target feature.

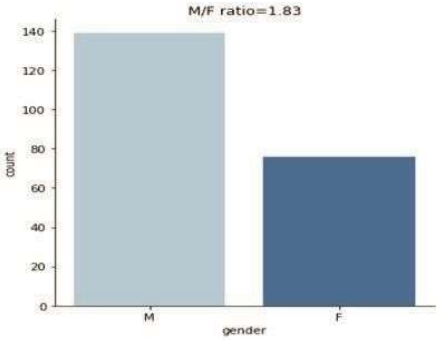

Fig. 2.
M/F ratio

Here, male : female ratio for one batch of students isapproximately equal to 2. It means that there are 2 male candidates appearing for placement drives for every 1 female candidate.

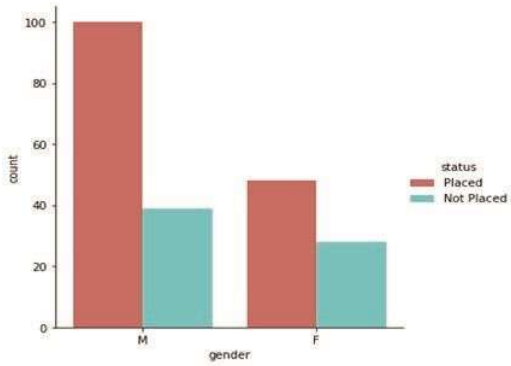

Fig. 3. Placement counts vs. gender

From the above graph it can be concluded that the count of placed male candidates in a batch is higher as compared to female candidates & the placement count is dependent on gender.

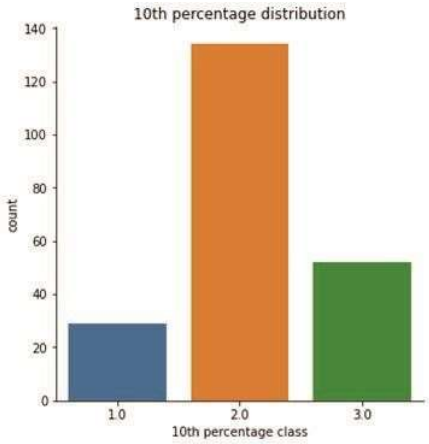

Fig. 4. 10th standard percentage distribution

In the above graph, class 1 represents students having scoresbetween 80-100%, class 2 represents students having scores between 60-80% and class 3 represents students having less than 60 % score in 10th standard.

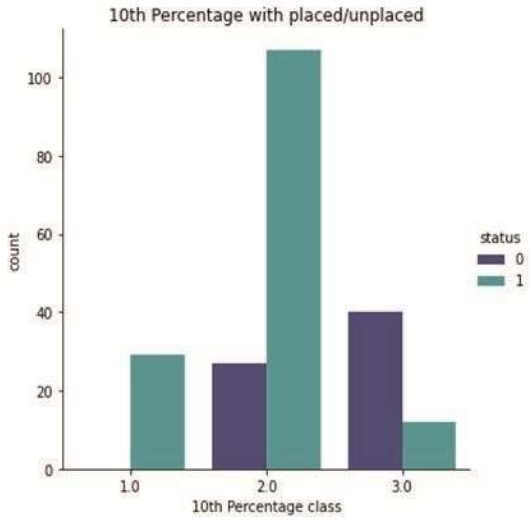

Fig. 5. Placement count vs. 10th percentage

From the above graph, it's observed that all the students having scores between 80-100% in 10th standard got placed.

Very few students having scores between 60-80% in 10th standard couldn't get placed. Whereas, most of the studentshaving below 60% score in 10th standard couldn't get placed.

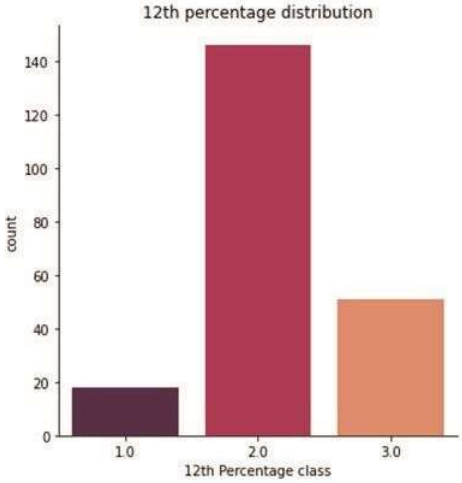

Fig. 6. 12th standard percentage distribution

In the above graph, class 1 represents students having scores between 80-100% , class 2 represents students having scores between 60-80% and class 3 represents students having less than 60 % score in 12th standard.

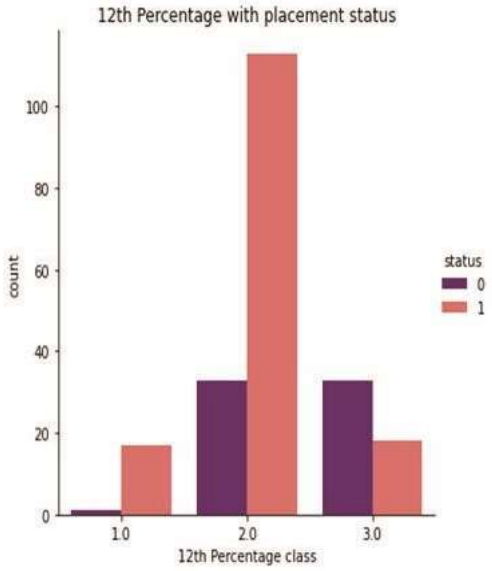

Fig. 7. Placement count vs. 12th percentage

From the above graph, it's observed that all the students having scores between 80-100% in 12th standard got placed. Very few students having scores between 60-80% in 12th standard couldn't get placed. Whereas, most of the students having below 60% score in 12th standard couldn't get placed.

G. Split Data:

Here the data is divided into two parts, i.e. training data and testing data. where 80% of the data is used to train our machine learning algorithm and the remaining 20% of thedata is used to test whether our trained machine learning model is performing well or not.

Machine learning algorithm:

- **Logistic regression:**

  Logistic regression is a statistical method used to determine the outcome of the dependent variable (y) based onthe values of the independent variable (x). In our problem, thedependent variable is the position state and the independent variables are the characteristics we selected in the previous step. This algorithm is mainly used for binary classification problems.[1]

- **Decision tree:**

  A decision tree is a graph like a tree in which the nodes represent the location from which we select the feature and ask the question, the edges represent the answer to the question; and the leaves represent the final output or label of the layer.[1]

- **Random Forest:**
  The Random Forest classifier consists of several decision trees applied to different subsets of our dataset, and the outputs of all decision trees are averaged to improve accuracy. accuracy of prediction.[1]

- **KNN:**
  K-NN stores all the training data into different classes based on the class labels and classifies new data by checking its similarity with data in the available classes.[1]

CONCLUSION

The problem of predicting campus locations can be solved using various machine learning algorithms such as Logistic Regression, Decision Trees, KNN, and Random Forests. Here, the logistic regression algorithm gives the highest accuracy of 95.34% for campus location prediction. Selected characteristics e.g. serial number, Gender, SSC Percentage, SSC Table . - Central/Other, HSC Percentage, HSC Specialization, Diploma Percentage, Diploma Specialization, Work Experience, E- Test Percentage and Salary Expectation lead to higher classification accuracy.

FUTURE SCOPE

Accuracy can be further increased through applying more advanced techniques such as deep learning and testing different neural network activation functions such as linear, sigmoid, tan h and ReLU.

We can also test different cross-validation techniques like 3-fold, 5-fold, 10-fold, 15-fold cross-validation to analyse the change in accuracy.

This model can be further improved according to the growing competition and can also be offered in a modifiable manner according to company specific criteria. It can then be added to the institute's website for students to check whether they are eligible for internship preparation or not.

REFERENCES

[1] Priyanka Shahane Department of Artificial Intelligence & Data Science, AISSMS Institute of Information Technology, Pune, Maharashtra, India priyankashahane04@gmail.com

[2] Ajay Shiv Sharma, Swaraj Prince, Shubham Kapoor, Keshav Kumar "PPS - Placement Prediction System using Logistic Regression" P 338,339 2014 IEEE International Conference on MOOC, Innovation and Technology in Education (MITE)

[3] Pothuganti Manvitha Department of Information Technology Sreenidhi Institute of Science and Technology, Neelam Swaroopa Department of Information Technology Sreenidhi Institute of Science and Technology

[4] Mrs. J. Samatha, Assistant Professor, Department of Computer Science and Engineering, Matrusri Engineering College, Saidabad, Hyderabad, Telangana D. Manjusha, B. Pooja, A. Usha, B.E Scholars, Department of Computer Science and Engineering, Matrusri Engineering College,Saidabad, Hyderabad, Telangana

[5]S. Elayidom, S. M. Idikkula, J. Alexander and A. Ojha, "Applying Data Mining Techniques for Placement Chance Prediction," 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009, pp. 669-671, doi: 10.1109/ACT.2009.169.

[6] J. Nagaria and S. V. S, "Utilizing Exploratory Data Analysis for the Prediction of Campus Placement for Educational Institutions," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225441.

[7] S. Venkatachalam, "Data Mining Classification and analytical model of prediction for Job Placements using Fuzzy Logic," 2021 IEEE International Conference on Trends in Electronics and Informatics (ICOEI), 2021.