

Expert System for Student Placement Prediction

Krishna Gandhi
Msc. Data Science and Big Data
Analytics
MIT- WPU
Kothrud,Pune, Maharashtra,
India.

Aadesh Dalvi
Msc. Data Science and Big Data
Analytics
MIT- WPU
Kothrud, Pune, Maharashtra,
India.

Aniket Walse
Msc. Data Science and Big Data
Analytics
MIT- WPU
Kothrud, Pune, Maharashtra,
India

Sachin Bhoite
Msc. Data Science and Big Data
Analytics
MIT- WPU
Kothrud, Pune, Maharashtra,
India

Abstract: - Data mining is a process of extracting and identifying previously unknown and potentially useful information or pattern from large amount of data using different methods and techniques. Data mining in a domain of education is known as Educational data mining (EDM). This paper discusses about an expertise system which can used as student placement prediction system. A statistical model is applied on a reputed college's past data after data pre-processing and feature selection. This model can be used to predict the percentage of chances of a student getting selected in campus placement. It will help students evaluating themselves and identifying which skills are essential.

Keywords:- Data Mining, Educational Data Mining, Expertise system, Statistical model, Data pre-processing, Feature selection.

1. INTRODUCTION

This model is about concerning those students who wants to get a better placement for better future. Sometimes it happens that student gets sidetracked from studies in initial Semesters and later they realize the importance of marks/CGPA. Basically, this will help them to enhance their performance and will make them believe that they can achieve their dream job.

Application of EDM is an evolving trend in the worldwide ^[1]. This will help college faculty to show the precise roadmap to students when it comes to placement and choose their career path.

It will guide colleges and institutions to maintain their reputation by making most of the placement. It drives students to ask questions regarding what can nurture them. It can give an overview to

Junior college students while selecting their stream.

- 1) What subjects to target?
- 2) What skills to improvise on?
- 3) Probability of getting placed after choosing their specialization?

Data visualization will help College students to get a clearer view regarding which stream they should choose. This can be done using different libraries of Python like Matplotlib, where student and faculties can visualize overview of each stream.

This paper describes the model to predict the percentage of skills required by engineering students pursuing Bachelors and Master's Degree with respect to company's skillset necessity. According to the rules generated, percentage of selection of students will vary. These rules are generated with the help of Domain Expert.

Percentage of Selection = (Criteria's Satisfied/Number of Criteria's) * 100

1.1 Literature Review

The researchers have studied several related national & international research papers, thesis to understand aims, technique used, various expert systems, datasets, data preprocessing approaches, features selection methods, etc.

Siddu P. Algur, Prashant Bhat and Nitin Kulkarni used two algorithms- Random Tree and J48 to construct a classification models using Decision Tree concept. The Random Tree classification model is more effective as compared to J48 classification model^[2].

Machine learning algorithms are applied in weka environment and R studio by K. Sreenivasa Rao, N. Swapna and P. Praveen Kumar. Results is tabulated and analyzed, It shows random tree algorithm gives 100% accuracy in prediction on their dataset and also in R environment Recursive Partitioning & Regression Tree performs better and gives 90% accuracy. We also accept that performance depends on nature of dataset^[3].

V.Ramesh, P. Parkavi and P. Yasodha also proved that Multilayer Perception algorithm is most

suitable for predicting student performance. MLP gives 87% prediction which is comparatively higher than other algorithms^[4].

K. Sripath Roy 1, K. Roopkanth, V. Uday Teja, V. Bhavana, J. Priyanka. The data is trained and tested with all three algorithms and out of all SVM gave more accuracy with 90.3% and then the XG Boost with 88.33% accuracy^[5].

Ajay Kumar Pal, met with his goal and proved that the top algorithm is Naïve Bayes Classification with an accuracy of 86.15% with an error average of 0.28 with others. He also conveyed that naïve Bayes has the potential to classify conventional methods^[6].

Sudheep Elayidom, Dr. Suman Mary Idikkula and Joseph Alexander, studying past data and and following the trend, and based on that the judgment for future will be given.

2. DATASET DESCRIPTION

The data used in this model is supplied by a well-known Engineering College situated in Pune, Maharashtra. Data generated is collected from the details given by graduates, post graduates, diploma holders in engineering of various streams during the year 2019. It includes students 10th, 12th or Diploma and semester-wise aggregation for Bachelors and Master's. Dataset contains 2330 tuples and 81 attributes holding multiple streamwise data of the students.

2.1 Data Pre-Processing

Data has redundant, incomplete, inconsistent and inaccurate entries. We discovered that there were many different attributes which seems to be superfluous and which won't affect our results. By consulting our Domain Expert, we decided to remove those attributes as well as tuples using tools like excel.

Entries with human errors seem to be illusory. So as per discussion with Expert we decide to apply mean to the data using Python.

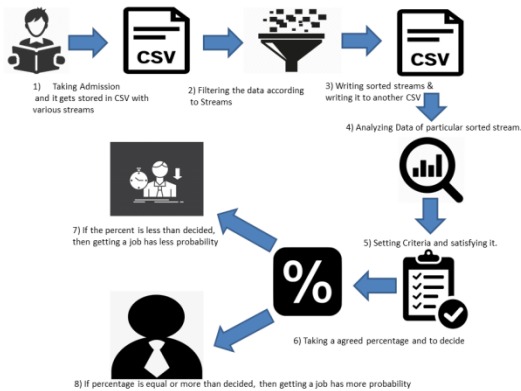
2.2 Feature Selection

Attributes impacting the placements of students were taken into consideration with the help of Expert advice.

Factors like 10th, 12th or Diploma and Degree Aggregation are too affecting the placement predictions for students as well as their non-educational attributes like work-experience, projects and external certification were also taken into consideration.

3. PROPOSED EXPERT SYSTEM

After analyzing a lot we came to an idea of creating our own system, we will help us to understand the genuine accuracies for placement.



- 1) A student fills the form and gets stored in CSV.
- 2) CSV contains all details of all the streams data of all students
- 3) Then system filters all streams and writes one by one detail in each new CSV.
- 4) Then domain expert analyzes it on each stream and sets criteria for all the attributes.
- 5) The system checks whether students get satisfied or not according to the specific criteria set.
- 6) Then the criteria will decide which students are going to be placed in a campus placement and which are not.

3.1 Pseudo Code Of The System

Pseudo is basically the demo code which we have decided to implement on our model. It is type of rule setting methodology where student have to fulfill all criteria's for getting a reputed placements.

1. Load all student data.

2. Form rule based on stream with the help of expert
3. Enter the data of new students for the prediction of the placement.
4. Calculate Number of criteria satisfied by that student.

Flag=0

If rule 1 satisfied

Flag=flag+1

Else

Flag=flag

If rule 2 satisfied

Flag=Flag+1

Else

Flag=flag

5. Perform this operation for all the criteria created.
6. Calculate our prediction by

Prediction percentage= (Flag/Number of rules) *100

4. EXPLORATORY ANALYSIS

A pie chart is a circular statistical illustration, which is divided into different parts to demonstrate numerical proportion.

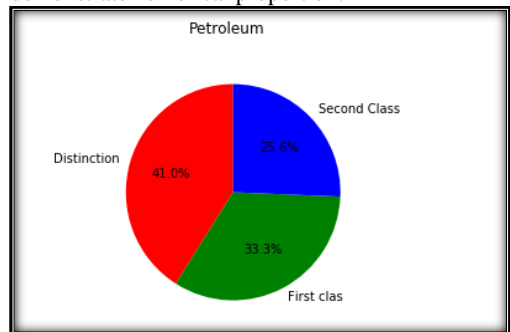


Figure.3.1 Pie chart of Class wise placement for Petroleum

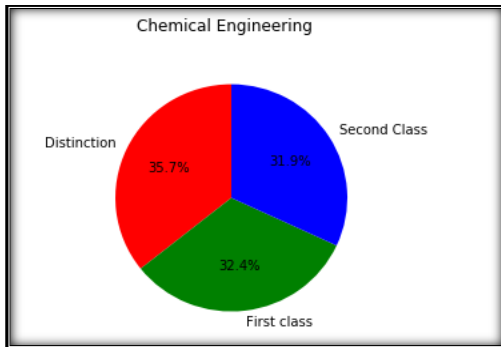


Figure.3.2 Pie chart of Class wise placement for Chemical Engineering

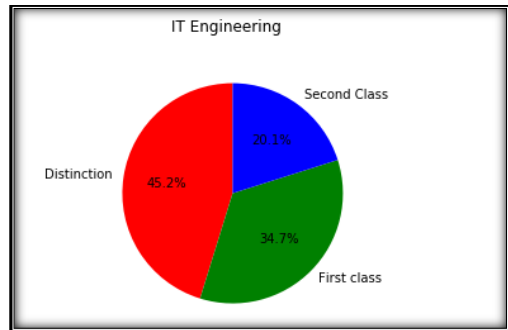


Figure.3.5 Pie chart of Class wise placement for IT Engineering

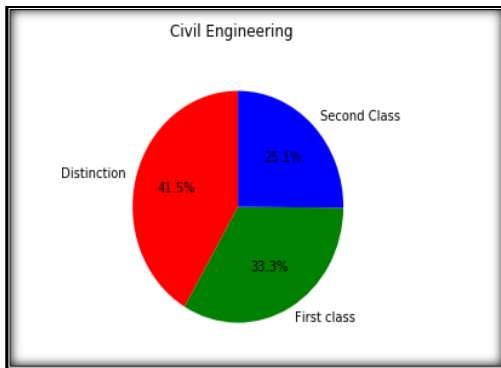


Figure.3.3 Pie chart of Class wise placement for Civil Engineering

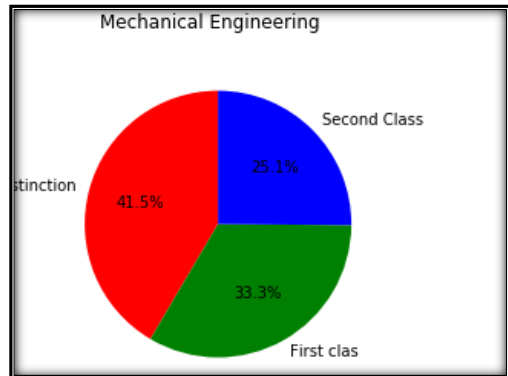


Figure.3.6 Pie chart of Class wise placement for Mechanical Engineering

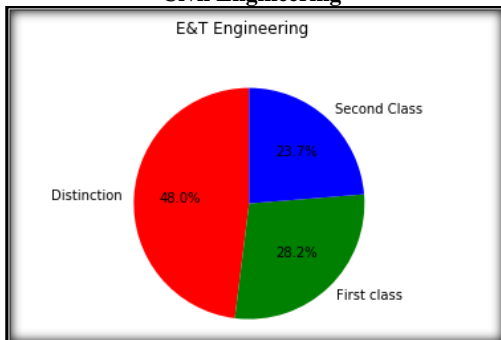


Figure.3.4 Pie chart of Class wise placement for E&TC Engineering

The pie charts shown above give the information about students placed in campus interview based on their grade as per each stream. This gives an overview to students about stream and importance of aggregates according to the stream.

5. CONCLUSION

This paper examines application of Educational Data Mining (EDM). This paper elaborates the model to create awareness for students to create a better careeristic pathway for their future. Students with the help of their professors and placement team can make use of this model to get better placement opportunities and enhance their skillsets.

In Future this model can be compared with existing Machine Learning Algorithms like Linear Regression, Logistic Regression and Decision tree which will help us to understand the accuracy of

percentage of Machine Learning and Statistics. We will come to know what accurate percentage to rely on with the comparison of statistics and Machine Learning.

Trends in Engineering, Vol. 1, No. 1, May 2009.

6. REFERENCES

- [1] Dr. Mohd Maqsood Ali,” ROLE OF DATA MINING IN EDUCATION SECTOR”, IJCSMC, Vol. 2, Issue. 4, April 2013, pg.374 – 383
- [2] Siddu P. Algur, Prashant Bhat and Nitin Kulkarni,“EDUCATIONAL DATA MINING: CLASSIFICATION TECHNIQUES FOR RECRUITMENT ANALYSIS” I.J. Modern Education and Computer Science, 2016, 2, 59-65
- [3] K. Sreenivasa Rao, N. Swapna, P. Praveen Kumar “EDUCATIONAL DATA MINING FOR STUDENT PLACEMENT PREDICTION USING MACHINE LEARNING ALGORITHMS”, International Journal of Engineering & Technology, 7 (1.2) (2018) 43-46
- [4] V. Ramesh, P. Parkavi and P. Yasodha,” PERFORMANCE ANALYSIS OF DATA MINING TECHNIQUES FOR PLACEMENT CHANCE PREDICTION”, International Journal of Scientific & Engineering Research Volume 2, Issue 8, August-2011.
- [5] K. Sripath Roy 1, K. Roopkanth, V. Uday Teja, V. Bhavana, J. Priyanka,” STUDENT CAREER PREDICTION USING ADVANCED MACHINE LEARNING TECHNIQUES”, International Journal of Engineering & Technology, 7 (2.20) (2018) 26-29.
- [6] Ajay Kumar Pal, “Classification Model of Prediction for Placement of Students”, Published Online November 2013 in MECS (<http://www.mecs-press.org/>),DOI: 10.5815/ijmecs.2013.11.07.
- [7] Sudheep Elayidom, Dr. Suman Mary Idikkula and Joseph Alexander, “Applying Data Mining using Statistical Technique for Career Selection”, International journal of Research