STUDENT DATA ANALYSIS

Selvakumar. G[1], Santosh Kumar. S[2], Keerthana.V[3], Sudarsana. R[4]

[1]*Associate professor, Sri Shakthi Institute of Engineering and Technology,*
[2]*Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology,*
[3]*Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology,*
[4]*Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology.*

Abstract— The Educational Institution's growth is directly related to the placement activities and the admission of the students. In order to increase the standard of the institution, an analysis is performed on the students' data and the visualization is obtained from which meaningful insights are produced. These insights help to identify the necessary steps that are to be taken in order to take the students and the institution to the next level. Insights are obtained for placement and admission process of the students.

Keywords— admissions, analysis, insights, placement activities, students' data, visualization

I. INTRODUCTION

Data analytics is the technique which explores the hidden information in the datasets and converts them into knowledge from which conclusions can be drawn for Business decisions. While the presentation of data in a pictorial or graphical format is Data Visualization. Visualization enables decision makers to grasp difficult concepts or identify new patterns through visualizing analytics. With interactive visualization, we can get interactive charts and graphs for a more detailed view, and also change what data we see and how it's processed. In educational Institution, the administrators have the access to the data sets of the students which can be of any format. These datasets are obtained and Analysis is performed in order to extract knowledge from the data. The knowledge obtained helps the institution to grow in admission and placement activities.

II. RELATED WORK

In this paper [1], the Data Mining algorithm ID3 was used to analyse the performance of the students in placement. The existing system was not so efficient, as the data was distributed among various instructors and this made it be difficult for collecting the scattered data for maintenance and analysis. Hence, to overcome the problem in the existing system the new system was proposed which is a web application that can be accessed by all the students or staffs and others. This also makes the micro level decision and also produces outcomes based on the queries given by the administration team. It gives an immediate analysis of the data which provides a suggestion that helps the college teaching and placement staff in changing their teaching and training policies or adding various new additional programs which would help the student to develop the required skills [4].

The main objective of the institution is to provide quality education and training for the students for better placement opportunity. In this paper [2], the analysis for the placement was done using decision tree algorithms like ID3, CHAID, and C4.5. This was implemented using Rapid Miner tool. Initially, the placement data of the passed out students was gathered and data pre-processing was done as the next phase, where the required attributes for prediction were identified and then analysed. . From the results, it was found that ID3 algorithm is appropriate for predicting student placement. ID3 gives 95.33% prediction which is higher than C4.5 and CHAID algorithm.

This paper [3] presents a comprehensive statistical method to identify the number of students those who are ready for placements and students those who are not fulfilling the basic criteria for placement from a large database of all computer engineering students of a college containing their academic record. WEKA tool was used to classify the students into clusters based on their performance. The important parameters are involved in identifying the students' skill in all the activities taken part. At the end of the analysis, it concludes how many students in the college are ready to fulfil all the basic requirements for taking part in a placement drive, how many students to improve their performance and how many students need to face difficulty in completing their degree. The final result demonstrates the student performance in academics, their eligibility to attend the placement drive and their inability to complete the course on time.

This paper [4] provides a survey of the existing work done on performance analysis and monitoring, and in understanding the Algorithm used in the existing system. The project aims at displaying the students' performance in the form of graph and score card on the single click of the user to enter the user's name and the unique ID, thus introducing automation and reducing efforts of staff in analysing student performance manually.

These papers suggest us to evaluate the students' data to identify their performance and it also suggests various algorithms that could be used for analysing the students' data efficiently. This kind of analysis reduces the work of the staffs and management in understanding the students' mentality and their performance.

III. SCOPE OF THE WORK

This system is mainly used to take the institute and its students to the next level by analysing the reason behind the students' choice in selecting that particular institute for their studies and their mind set regarding the placements. The result of this analysis will be comprehensive that gives all the possible suggestions and recommendations in improving the procedures followed by the institute for its betterment. Various visualization techniques are used to represent the result of this analysis for better and easy understanding. This project can also be extended to various scenarios in business to improve their strategies followed.

IV. IMPLEMENTATION

A. *Data Collection*

Data collection is the first step that is carried out in the process. Here, we collect all the data regarding the admission of each student to the institute for each academic year and about the placement of the final students in each department. Once the required data is collected it is all put into either in the form of excel file or csv file then the analysis and visualization is done using RStudio using which meaningful insights are obtained.

B. *Importing datasets and Deciding on variables*

The data sets which are in the form of excel and csv are imported into the RStudio. We have to decide the variables on which the visualization should be done. The variable selection differs for each visualisation based on the datasets. These variables are the curial factors that bring out the meaningful result.

C. *Visualisation using R*

R Studio is a statistical tool that identifies meaning in the data. The packages such as dplyr, ggplo2, plotly are used in order to perform visualisation on the data that provide insights for the Institution.

D. *Obtaining insights*

Based on the visualisation insights can be derived which provides recommendations, knowledge for the institution. These insights provide a clear path in which the institution must move on and also provides the current status of the institution in placement and admission. Thus it helps the institution to improvise the standard.

E. *Displaying insights*

The insights obtained for various visualisation is viewed through a static web page. According to the need, insights can be viewed in this web page.

V. EXPERIMENT RESULTS

Visualization was performed on the students' data and insights were obtained. The insights obtained mainly focuses on placement and admission activities. Some on them are given below.
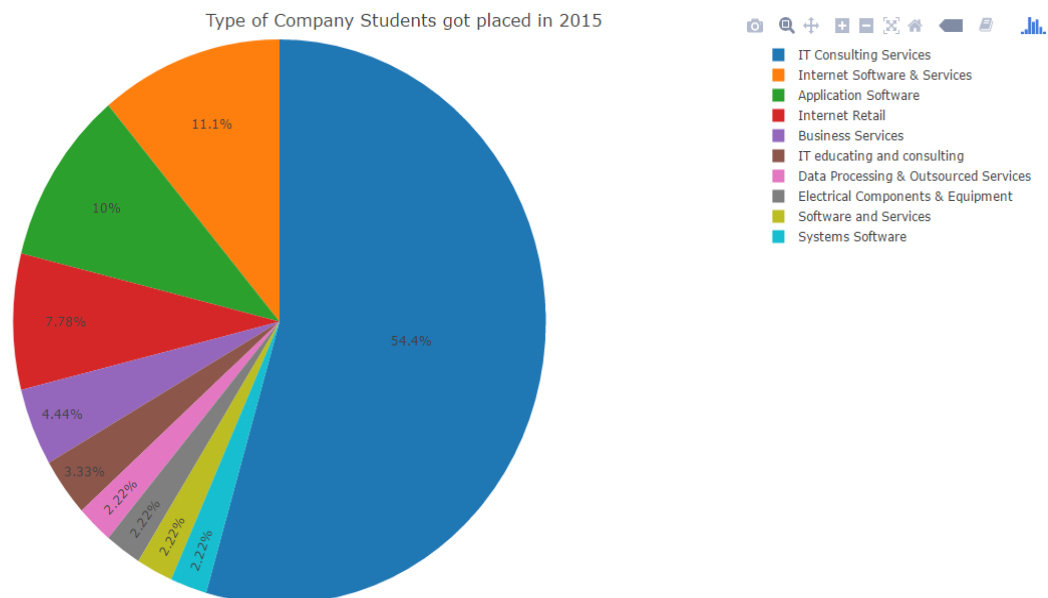
Company type in 2015

Figure 1

This shows that majority of the students get placed in the IT Consulting Service companies. The skills of the students must be developed to compete for the companies in which the standard of the recruitment process is high. The companies in the domains: system software, software and services, electrical components and equipment, data processing and outsourced services has a low recruitment count as the expectations were high. Hence additional training has to be given for the students in these domains.

Promotions in each district
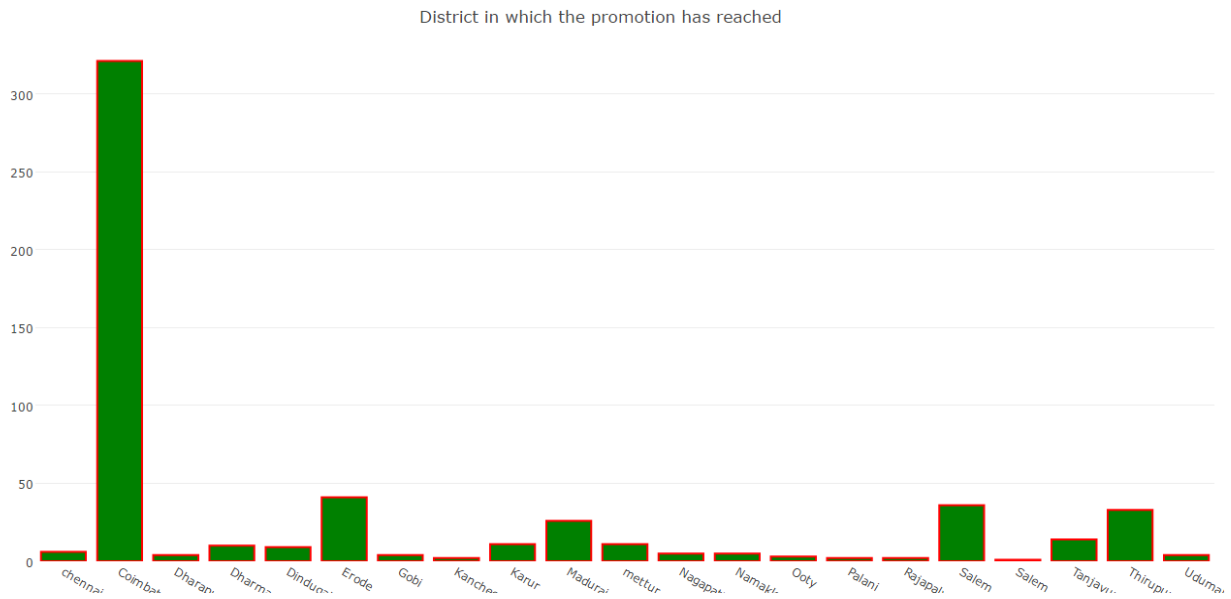
Figure 2

The promotions made by the college has reached all these districts. Since the college is located in Coimbatore maximum number of admissions is from this place. The promotion has reached the other districts such as Erode, Salem, Thirupur and Madurai in a good manner. If the promotions like educational fair, paper ads, promotion at schools are increased the admission will be improvised.

The overall view on promotion is given below,

Figure 3

This shows that on the whole, promotions about the college in various schools and educational fair should be increased.
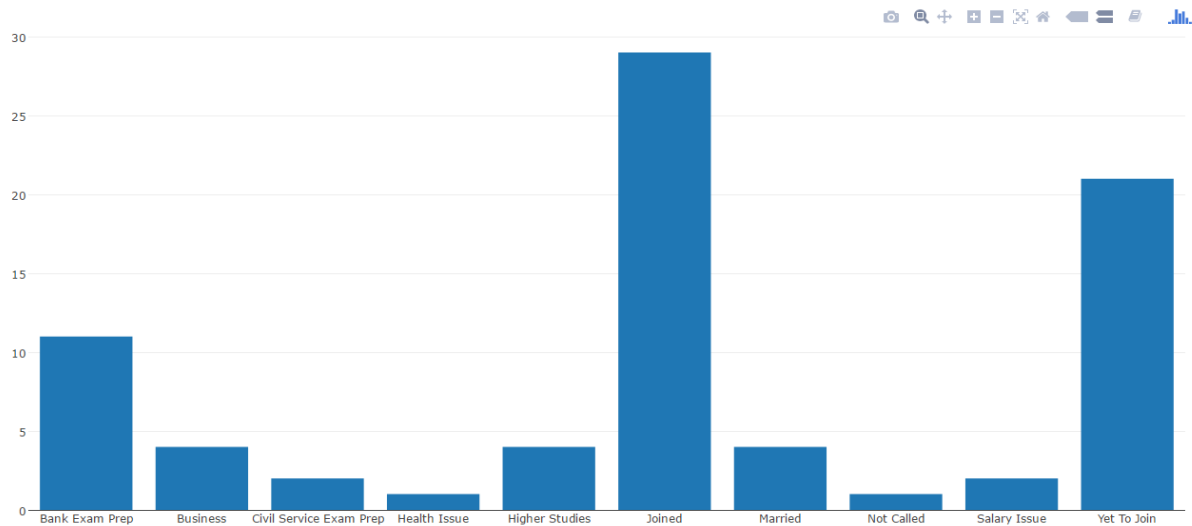
Students' career status during graduation

Figure 4

All the students who got placed in the companies have chosen their career in various fields. Few haven't joined due to personal reasons; few haven't because of the other career options. Almost 35% haven't joined the company in which they were placed in. this may create a bad impression on the institution. Hence students will have to be separated according to the career path which they choose. If the companies have not called the recruited students' then these companies should be avoided for the next academic placement process.

VI. CONCLUSION

The system is mainly used to analyze the student's data more deeply and visualize the datasets in order to provide meaningful knowledge of the dataset. The visualization is user interactive that can produce the outcome for all the data that is given by the user. It guides the institution to take necessary steps to improve the standard and thus increase the admission count.

REFERENCES

[1] Siddhi Parekh, Ankit Parekh, Ameya Nadkarni, and Riya Mehta, "*Results and Placement Analysis and Prediction using Data Mining and Dashboard*", Volume 137 – No.13, March 2016.

[2] T. Jeevalatha, N. Ananthi, and D. Saravana Kumar," Performance Analysis of Undergraduate Students Placement Selection using Decision Tree Algorithms", Volume 108 – No 15, December 2014.

[3] Praveen Rani, and Dr. Rajan Vohra*," Generating Placement Intelligence in Higher Education Using Data Mining ",* Vol.6(3),2015,2298-2302.

[4] Snehal Kekane, Dipika Khairnar, Rohini Patil, Prof. S. R. Vispute, Prof. N. Gawande, "*Automatic Student Performance Analysis and Monitoring",* Vol. 4, Issue 1, January 2016.