

Application of Data mining in Automation of Placement Cell

Ashish Zunjarrao¹, Tushar Waghela², Prasad Shetty³, Kavita Bathe⁴

1 Student, Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Maharashtra, India

2 Student, Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Maharashtra, India

3 Student, Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Maharashtra, India

4 Assistant Professor, Computer Engineering, K.J. Somaiya Institute of Engineering and Information Technology, Maharashtra, India

Abstract- Now a days competition has increased to a higher level and the need is to built-up the smartness along with accuracy and practical conduct. The Placement cell and Alumni Information System (along with data mining techniques like clustering and classification), formulated by thorough study, will help to generate various reports for forecasting about the number of students that might get placed via recruitment process. The details of the students will be extracted from the centralized database and be provided to the companies for the process of recruitment. This will help the organization's students to compete efficiently in recruitment process.

concern officer of organization for attending the campus interviews. This process is maintained manually, like maintenance of their resume in papers or maintaining records in excel sheets. This can be automated by designing web based software.

The proposed system will measure various parameters to understand the student's desire for certain company profile and ability to get placed. It will help the colleges to overcome the difficulty in keeping records of thousands of students as well as searching for a student eligible for recruitment criteria. This will thereby provide better communication facility between TPO and students. The Alumni information System will also keep records of all the pass out students for future use thereby making the post recruitment process clear.

Key Words: Recruitment, alumni, classification, clustering, centralized database, data mining.

1.INTRODUCTION

Now a days campus placements are conducted in many colleges. Various software as well as other sector companies are conducting campus selection drives for selecting merit candidates. For this students should provide their curriculum vitae to the

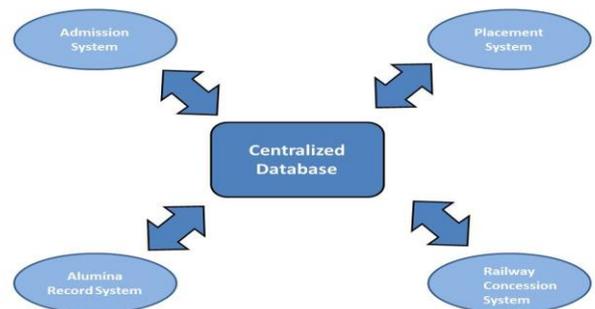


Fig. 1. Centralized DBMS

The system contains all the information about the students which is stored initially at the time of admission. It contains thousands of records of the students containing various details that are not in use and are waste which can be used to provide various functions such as automated CV generation, annual reports generation, etc.

Various Data mining algorithms will be applied on the data that have been previously collected by the organization for the training and placement purpose to find out meaningful pattern and generate reports from the database for further interpretation which will be helpful for the organization in making future decisions

to improve process to predict the number of students that might get placed among the number of students which are appearing for the recruitment process.

The proposed system will also provide automated and efficient communication facility between students and TPO for better communication. Records of pass out students are being maintained by the system for future use.

2. EXISTING SYSTEM

In existing System student has to fill data for repeatedly during the entire process. Each time the database is stored in the System creating the replicas of same data in the system and communication between recruiters, students and TPO is managed manually. This is time consuming as well as possibility of human error with lots of human efforts for communication with individual. In the existing system there is lot of manual workload on the TPO of the institute. The system only collect the data from the students, the validation and verification of the data needs to be done manually. It has following limitations each time data needed be verified which results in wastage of time, the sorting of candidate according to recruiters criteria is needed to be done manually, there is no automated system for communication hence all the communication has to be managed by TPO and there is no special facility available in the system to keep track of pass out students.

3. PROPOSED SYSTEM

The proposed system is motivated by the idea to provide best utilization of the available resources i.e. utilizing a single database for all the processes. It overcomes various limitations of existing system making our system suitable for its purpose. Our project focuses on making the placement procedure simple with effective database management and automated communication which saves the efforts and time of both TPO and students with the help of using data mining algorithms especially clustering algorithm.

4. METHDOLOGY

Data mining is a technology which helps organizations to process the data through algorithms to study and analyze meaningful patters and correlations from large databases. Functions of the data mining that will be used are clustering and classification techniques.

Classification data mining technique will be used for finding models that will analyze and classify data items into several predefined classes according to some specified constraints and clustering technique will identify the finite set of clusters to describe the data. Data mining algorithms will be applied on the existing database which will generate graphical reports. This algorithm will be used for the generalization of the data according to the different instances. This technique will be implemented using data mining software tools like Weka tool.

Naive Bayesian and ID3-decision tree learning classification algorithms will be used. Naive Bayesian will be used which is a supervised learning method which is based on applying Baye's theorem providing a statistical method for the classification which provides the proper prediction of the future by combining the prior knowledge the observed data.

Formulae used

$$P(y | x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n | y)}{P(x_1, \dots, x_n)}$$

ID3 is an algorithm that was invented by Ross Quinlan which is used to generate a decision tree from dataset and tells us that to which class elements of dataset belongs to.

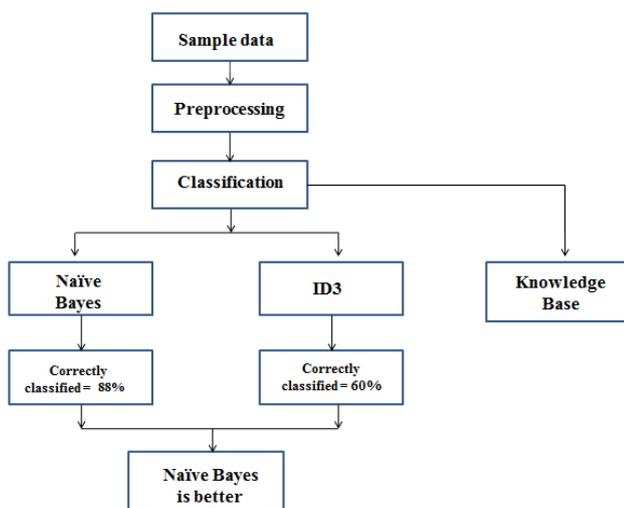
Formulae used

1. Entropy

$$H(S) = - \sum_{x \in X} p(x) \log_2 p(x)$$

2. Information Gain

$$IG(A, S) = H(S) - \sum_{t \in T} p(t)H(t)$$



The above flow chat depicts that we have collected the sample dataset from the institute and preprocessing on the data is done. Classification techniques like Naïve Bayes and ID3 algorithm are used to predict the output of the candidates being placed in upcoming placement procedure.

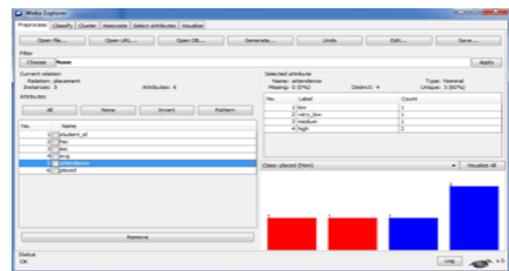
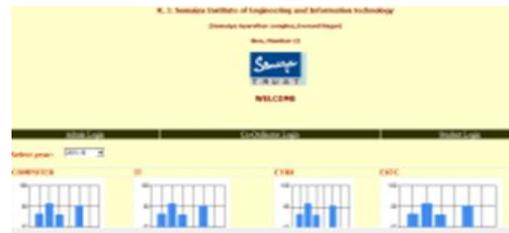
5. RESULT AND IMPLEMENTATION

We had collected dataset containing 200 samples . Using WEKA tool we had applied above mentioned algorithms i.e Naïve Bayes and ID3 and compared the results analyze which algorithm gives better prediction about the candidate being placed for the recruitment process.

The result obtained were as follows :

Correctly Classified using Naïve Bayes = 88%

Correctly Classified using ID3 = 60%



From the above mentioned results it is clear that the Naïve bayes algorithm will classify the given data set elements more correctly than the ID3 and it will be helpful for predicting future results.

REFERENCES

[1]An Interactive Online Training and Placement System
International Journal of Advanced Research in Computer Science and Software Engineering. Volume 3, Issue 12, December 2013.

[2]Automated Extraction of Database Interactions in Web Applications Minh Ngoc Ngo; Hee Beng Kuan Tan; Trinh, D. Program Comprehension, 2006. ICPC 2006. 14th IEEE International Conference on Year: 2006

[3]Wikipedia, en.wikipedia.org

[4] <https://www.somaiya.edu/kjsieit>

[5] <http://www.ibm.com/developerworks/library/ba-data-mining-techniques/>