

A Survey Paper For Managing Training & Placement Data Efficiently Online Through Data Mining Techniques



#¹Chaitrali Raut, #²Dipali Kakade, #³Kishor Chindhe, #⁴Surajsingh Badgujra,
#⁵Prof. B.Padmavathi

¹chaitraliraut94@gmail.com
²dipalikakade555@gmail.com
³kishor.chinde@gmail.com
⁴jasvndr03singh@gmail.com
⁵b.padmavathi@raisoni.net

#¹²³⁴Department of Computer Engineering
#⁵Prof. Department of Computer Engineering
GHRCEM,Pune.

ABSTRACT

A web portal will be designed for training and placement data which can be used by placement cell of a college. The objective is to design a system that provides functionalities to perform the activities related to placement services belonging to the institute i.e. college. It is completely modular architecture system. The modular architecture helps to add or replace or remove certain modules in future as per the institute needs. The system can be further used as an application for the TPO of the college who manages the student data regarding to training and placement. The system provides opportunities to the student community to use collective intelligence to increase selection ratio and eases out process of creation of management information automatically. The admin of the system will have the right to manage all the things. System Admin is divided into three categories.

Keywords: TPO, TnP, Security, Authorizing, Forum.

ARTICLE INFO

Article History

Received: 9th February 2016

Received in revised form :

9th February 2016

Accepted: 12th February 2016

Published online :

13th February 2016

I. INTRODUCTION

Using user name and password students in different colleges in RGI can fill data sitting anywhere but should have access to internet. Admin will type the query and he will get the necessary output. Members of the college can analyze the data in the form of graph representation of the students criteria and another activities. Students can also fill the data offline using intranet. This portal is develop in PHP language which is platform independent and open source. It is basically develop for college purpose but it can be extended to a large institute by modifying the software as per institution needs. Modification can be done only by the Administrator which are divided into three and given different priorities.

This portal can be used by students to fill data sitting anywhere and anytime. Links and information /details/result about documentation on TPO cell manager is hidden. The training and placement engages in consistent industry

interaction by liaising with various industrial establishment and corporate houses for summer internships, placement quest talks and visits. It acts as an interface among students alumni and the employment community. It bridges the gap between industry and academics. It also provides placement opportunity for passed out students by informing them about the arriving company and its respective criteria. If the passed out alumni fits in the respective company criteria, then he has a chance to place through college campus.

The software provides a data panel for every student to fill his/her personal as well as educational details. These details are then analyze by using various algorithms such as k-means, apriori etc. Students in the college are clustered accordingly. The clustered students are represented with the help of bar graphs. Thus users can easily grasp the type of students in the college campus. The knowledge grasped from these graphs can then be used by the administrator of

the college to take collective actions and help students in getting placed. Different workshops can then be conducted to improve the knowledge of students regarding placement process. In short these software also helps in increasing the reputation of college.

II. LITERATURE REVIEW

In traditional systems data was collected from each department by respective coordinator and then send to central office for further processing. Every year these process was repeated for new batch of students. It use to take lot of time and data was erroneous. Also data was not all secure as it travelled a long path of people. When results of students were out, again the coordinator was responsible for modifying the data and sending it to central office. These was really a tedious task. Students were also informed late about the arriving company and they were not able to prepare fully.

The new software is going to be better than traditional software as it reduces the manual work and collects data efficiently. It automatically provides access to its user by entering their respective username and password. It inform students about the latest updates through mail and it is very secure. The notice about any company is also displayed on the panel per student if he fits in the company criteria. It uses PHP which is open source. K-means clustering algorithm is used for categorizing student's personal and educational details.

III. EDUCATIONAL DATA MINING

Data mining (DM) is a series of data analysis techniques applied to extract hidden knowledge from server log data by performing two major tasks: Pattern discovery and predictive modeling. Educational data mining (EDM) is a field which adopts data mining algorithms to solve educational issues. Following issues are solved by clustering the data in to structured form:

- Enhance the decision processes in higher learning institutions.
- Streamline efficiency in the decision making process.
- Achieve specific objectives.
- Suggest certain courses that might be valuable for each class of learners.
- Find the most course effective way of improving retention and grades.
- Select the most qualified applicants for graduation.
- Help to admit students who will do well in higher education settings.
- Based on the theory of bounded rationality, decision-making is a fully rational process of finding an optimal choice given the information available . An ideal program evaluation framework should provide multiple facets of information to decision makers. Therefore, integrating more than

one data source and analytic method is essential for an effective program evaluation.

Engagement level:

Engagement is considered to be a key variable for enabling and encouraging learners to interact with the material ,with the instructor, and with one another, as well as for learning in general. In this study, engagement level was measured by the frequency of various learning interactions that happened within the LMS. Variables under the category "Student Engagement Variable" in Table 1 were applied to measure each student's engagement level ,which included:

- Average frequency of login courses.
- Average frequency of clicks per course.
- Students having aggregate as distinction.
- Students having aggregate as first class.
- Students having aggregate as higher second class.
- Students having aggregate as second class.
- Students having pass class.
- Each Student Permanent Registration Number (PRN).
- Student's sports and other curricular activities.

IV. METHODOLOGIES

The Apriori algorithm:

Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent itemsets determine by apriori can be used to determine association rule which highlight general trends in the database: this has application in domain such as Market Basket Analysis.

K-means algorithm:

k-means clustering is a method of vector quantization ,originally from signal processing, that is popular for cluster analysis in data mining .K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean , surving as a prototype of the cluster .These results in a partition of the data space Voronoi cells.

System Architecture:

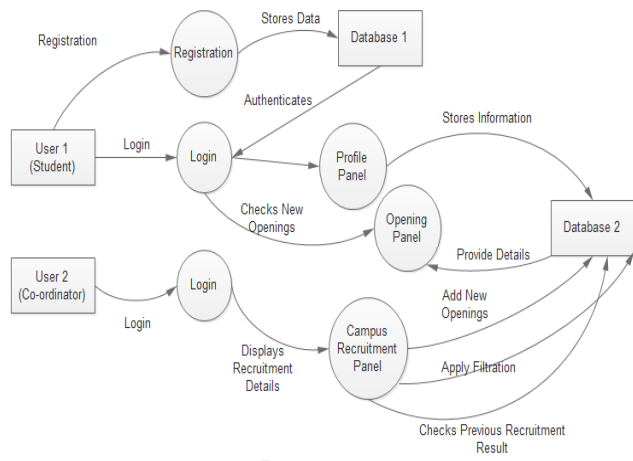


Fig 1. System Architecture

V. RESULT

Student clustering analysis

K-means algorithm was applied to group of student’s based on their percentage. (Internal Standardization = Range; Maximum Number of Clusters = 6). Total clusters were limited to avoid trivially small or exclusive groups, the identification of which was outside the purposes of this case study. A pass rate equal to “1” means a student passed all courses during the period of analysis. A pass rate equal to “0” means a student failed all courses during the period of analysis. A pass rate between “0” and “1” means a student passed some, but not all, courses during the period of analysis. In clustering analysis, pass rate was set up as the standard for classification and six clusters were generated.

- Cluster 1 (students, pass rate >50): Cluster 1 consists of students who are below 50 percent aggregate.
- Cluster 2 (students, pass rate = <50 &&>55): Students with 2nd class percent aggregate.
- Cluster 3 (students, pass rate =<55 &&<60): Students with higher 2nd class percent aggregate.
- Cluster 4 (students, pass rate = <60 &&<66): Students with 1st class percent aggregate.
- Cluster 5 (students, pass rate =60 &&>66): Students with distinction as aggregate.
- Cluster 6 (3,397 students with ATKT): Students with fails ATKT not display aggregate.

Variables CL1 CL2 CL3 CL4 CL5 CL6

Number of Students	316 320 594 601 2311 3397
Pass rate	0 0 0 594 601 0 0
second class	3188820 0 0 0 0
Higher second class	1 0 0 0 0 297
First class	0 320 0 97
Distinction	4 0 2311 0
Age	16.91 17.06 16.69 16.82 16.6 16.59
Grade_Avg	50.11 52.82 22.44 20.33
Course_Access_Avg	50.11 52.82 22.44 20.33
Course_Access_Avg	4.08 5.28 2.78 4.22 8.28 9.57
Login_Avg	29.4 24.35 23.58 19.18 47.92 46.42
Module_Access_Avg	156.18 145.02 882.7 102.38 249.16 888.79
No_Fail	156.18 145.02 882.7 102.48 349.17 666.49
No_Pass	1.52 1.7 0 0 1.59 1.44
No_course	2.76 3.03 1.88 1.39 1.59 1.88

Table 1. Result analysis

The clusters generated from cluster analysis were associated with two geographical variables: city and state, Findings below were summarized from the clustering analysis.

- 1) Students having Aggregate percentage as pass class.
- 2) Students having Aggregate percentage as second class
- 3) Students having Aggregate percentage as higher second class.
- 4) Students having Aggregate percentage as first
- 5) Students having Aggregate percentage as distinction.
- 6) Students having ATKT.

Putting different queries in my-SQL coordinator can get any data about students. It may be student having gap in the academic years or it might be their personal information regarding state, board of education etc.

VI.CONCLUSION

In the existing system, maximum work goes manually and it is error lying system , takes time for any changes in the system .We have developed a software which is error free, compact, flexible and which is very secure .The big problem is the searching; sorting and updating of the student data and no any notification method available for giving information to student expect the notice board. Therefore we have provided notice for students on their email-id as well as their OTnP accounts. Student automation system is web application which get HRs desired information without any delay. The student information which is stored in the database can be accessed any time by using SAS there will be no wastage of resources in any educational institution or organization.

REFERENCES

- [1] Kriegel H.P., Karsten M., Borgwardt, P.K., Pryakhin Alexey, Schubert M. and Zimek A.,” Future trends in data mining”, Data Min Knowl Disc., 2007.
- [2] I.Elsa, “Mining Simple and Complex patterns efficiently using binary decision diagram”, www.Biblioteca.Net, June 2009.
- [3] Bartolini I., Ciaccia P., Ntoutsis I, Patella M. and Theodoridis Y.,”A unified and Flexible Framework for comparing simple and complex patterns, Patterns for Next Generation Database System, 2004.
- [4] Han J., Cheng H., Xin D. and Yan X., “Frequent patterns: current status and future directions”, Data Min Know Disc, 2007
- [5] Frayyad U. and Piatetsky-Shapiro G. Smyth P,” Knowledge discovery and data mining: Towards a unifying framework. In: Proceedings of the 2nd ACM international conference on knowledge discovery and data mining(KDD),Portland, OR, pp 82-88, 1996.
- [6] Tzani G. and Berheridis C.,” Mining for Mutually Exclusive Items in Transaction Databases”, http://lpis.csd.auth.gr/publications/Tzani_IJDWM07.pdf
- [7]. Jing Luan, PhD Chief Planning and Research Officer, Cabrillo College Founder, Knowledge Discovery Laboratories “Data Mining Applications in Higher Education”.
- [8]. Juneja, Deepti, et al. "A novel approach to construct decision tree using quick C4.5 algorithm." Oriental Journal of Computer Science & Technology Vol. 3(2), 305-310 (2010) (2010): 6. Web. 18 Feb. 2013.
- [9]. [Online] http://www.creighton.edu/fileadmin/user/HSL/docs/ref/Searching__Recall_Precision.pdf
- [10]. [Online] <http://weka.wikispaces.com/Primer>.