Data Mining Model for Domain Selection

Miss. Amruta R. Jadhav[1], Miss. Rohini V. Pillai [2]

Student, Computer, NMIET, Pune, India [1, 2]

Abstract: Now a days in engineering colleges, domain selection process for project is not been focused seriously the manual procedure of selecting domain consumes unnecessarily too much time. (We can say around 2-3 months). Students in final year need to find the domain details on their own or another option is to ask the queries to respective teachers. It is really irritating as well as time consuming to tell every individual group to explain same domain. Here our system starts working. In our system there will be one module called student in that the sub module called domain will help students to know the details about the types of domain and the description of each and every domain in depth from the list. Moreover we are going to provide an aptitude test on domains, so that student's interest in particular domain can be calculated. In this module students need to add there percentage marks (transcripts). This data will be displayed to teachers in there login panel. They will be having the result of domain aptitude of every individual student as well as the list of students having domains in common. All these results will help teachers to do the grouping of students in particular domain. The teachers even can combine two domains which can co relate to each other so in result the students as well (e.g. Data mining, database, networking security). Coming back to current system TPO's in the colleges work is to get the campus details and they forward the same mail to all the students even though they are not eligible for a particular campus drive. Along with this TPO's are having hard copies of students' data individually. It is time consuming to sort out the eligible and non-eligible personalities. Our system will work smartly in this area as well. We will be providing dynamic categorization in which TPO's will be having the details of Students like academic marks and extra curriculum (technical, on-technical).TPO will work as a strong bridge between industries and students. As the data is available on web portal, TPO will forward the campus drive information only to the eligible students. Another module is there in which students can share study related material (PPT's, PDF's, audios, videos) with other students as well as with teachers.

Keywords: DataMining, Co-operative Learning, Domain Selection, C4.5

I. INTRODUCTION

Consistently, numerous building researchers need to arrange themselves for their major/final year project. The final year project plays an important role in demonstrating the effectiveness of studying results of modules that the students have taken through their studies. We are going to implement our own system which will help in domain selection process of students of final year project. The aptitude on the basis of domains will be conducted in student module.

The result of domain selection will be displayed to teacher module so that they can understand and divide the groups according to resulting domain. The another module called TPO will work as a bridge between students and companies.

It will help in finding the eligible students for particular campus drive in less time. The students which are eligible for particular campus drive will only know about the drive via mail.

Based on the results, several judgments are offered to upgrade student fulfilment. System is feasible in the point of accurately process the TPO Centers and for finding project topic in colleges. Methods we present a class of linear multi-regression models that are developed to create models that are demonstrate to each student and also take into explanation a considerable number of appearance that relate to a student past achievement, course components, and students commitment and achievement.

These models appraisal a small number of relapse models that are mutual across the different students along with student-specific linear sequence functions to expedite collection. Our preliminary assessment on a large set of students, courses, and activity shows that these models are capable of better the achievement.

In current oral system of project selection it is really very time consuming around months for both students and teachers to select the accurate academic project. It is still confusion for the students for the first time what the domain actually means.

Due to lack of knowledge about the domain, most students choose the domains randomly and fail to understand and work in the corresponding domain. Proper domain understanding and selection for academic BE projects the solution will be the resulting department according to test results for individual as well as group of students having common department after the test. The marks entered by the Students will be visible to the TPO who will return can send the data as well as mail to the students in the eligibility standards.

The list and description about particular domain gives clear idea to the students for selection process. Aptitude on domain selection will really heal the confusion and time wastage. Aptitude selection will lets students understand their real interests and will provide there suiting domain in which a particular student can at least do research by his/her point of interest.

We also added one more module in which students can share their knowledge, or information related to education with their Classmates as well as teachers TPO always need to send campus information to all students. It happens many times that some students are not eligible for specific criteria, still they ask about the drive to the Respecting TPO. So it is little bit difficult to handle no. of students for giving same Information all the time.

II. RELATED WORK

Describes the use of decision tree and rule inauguration in data mining applications of methods for allocation and relapse that have been developed in the range of pattern recognition, demography, and machine learning, these are of individual activity for data mining since they appropriate symbolic and intelligible representations. This facet is useful in the business and profitable applications [1].

This paper describes the use of decision tree and rule extracting in data mining applications of methods for classification and regression that have been developed in the fields of pattern recognition, statistics, and machine learn, these are of particular interest for data mining since they utilize symbolic and interpretable representations. These aspects are useful in the industrial and commercial applications [2].

The Bayesian network classifiers is related for anticipate the student's academic achievement and to achieve a model. This model helps to classify the drop outs and students who need special attention earlier and allow the teacher to supply correct advocate. In Addition, authentic indicator is useful in many various contexts. For example, identifying extraordinary students for scholarships and weak students who are likely to fails critical for apportion limited tutoring resources [3].

This paper specifies about what is data mining, Frequent Pattern Mining, Clustering, Classification , Probabilistic classification, Decision tree Classifier with example which is useful for understanding the knowledge of data mining[4].

Co-operative learning is used in teaching different issues on various educational levels-from fundamental to overhead. Objective of this paper is as follows; establish a separate learning alternative of Japanese college students. Study a variety of strategies that combine co-operative individual learning. [5].

Suggesting one way of inducing learners willingness to work with others, thus making co-operative learning successful, shifting many students towards their favouring co-operation the disadvantages in this paper is, only based on co-operative learning. It doesn't focuses on students capabilities and interests according to subjects in a particular domain area [7].

Based on a collaborative learn experience with hundreds of students over three consecutive years, that an approach using domain independent learning that is transferable to current e-learning plat- forms helps both students and teachers to manage student collaboration better.

The approach draws on a domain-independent modeling method of collaborative learning based on data mining that helps clarify which user model issues are to be considered. [8]

Regression analysis is used to examine the significance of four independent variables: cumulative grade point average prior to enrolling in intermediate accounting, grade in the introductory financial accounting class, grade in the introductory managerial accounting class, and score on a diagnostic assessment used to measure general financial accounting knowledge. Based on the results, several recommendations are offered to improve student performance. [10]
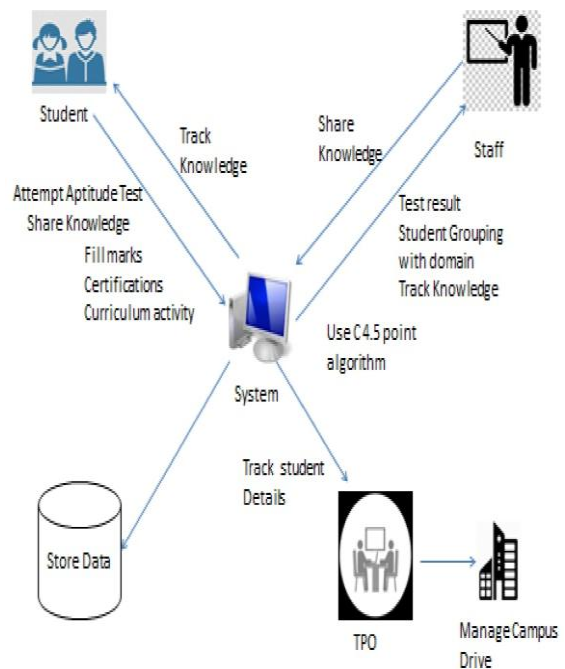
III. SYSTEM ARCHITECTURE

Fig. 1 Proposed System

As shown in above architecture our system is capable for calculating the student results according to their point of interest. Those results will be shared with the staff and

also stored into the database for further usage. Also every student user of this system is capable of sharing their knowledge by using share knowledge facility in our project. The very next module is the staffs who is able share knowledge as well as get test results of students also track the details of every student shared knowledge. Every staff can track which domain for project is suggested for which student for project making.

In another module TPO can get help from system for managing campus into the college. With the help of the system TPO will be able to watch the details of every student regarding student's academic criteria, certifications, curriculum activities, according this TPO will get classified student lists. So it is very helpful to manage campus in and out of colleges.

For all this our system is using some technical algorithm like c4.5 for data mining and also some searching and sorting algorithm as mentioned below.

A. C4.5:

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. This algorithm has a few base cases. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

All the samples in the list belong to the same class.

1. When this happens, it simply makes a leaf node for the decision tree saying to choose that class.

2. None of the features provide any information gain. In this case, C4.5 generates a decision node higher up the tree using the expected value of the class.

3. Instance of previously-unseen class encountered. Again, C4.5 makes a decision node higher up the tree using the expected value

C4.5 is implemented recursively with this following sequence

1. Check if algorithm satisfies termination criteria
2. Computer information-theoretic criteria for all attributes
3. Choose best attribute according to the information-theoretic criteria
4. Create a decision node based on the best attribute in step 3
5. Induce (i.e. split) the dataset based on newly created decision node in step 4
6. For all sub-dataset in step 5, call C4.5 algorithm to get a sub-tree (recursive call)
7. Attach the tree obtained in step 6 to the decision node in step 4
8. Return tree

B. For Searching:
Binary Search:
At the point when the qualities are in sorted request, a superior approach than the one given above is to utilize double hunt. The calculation for twofold inquiry begins by taking a gander at the center thing x. In the event that x is equivalent to v, it stops and returns genuine.

Else, it utilizes the relative requesting of x and v to take out portion of the exhibit (if v is not as much as x, then it can't be put away to one side of x in the exhibit; correspondingly, in the event that it is more noteworthy than x, it can't be put away to one side of x). When half of the exhibit has been killed, the calculation begins again by taking a gander at the center thing in the staying half. It stops when it finds v or when the whole cluster has been dispensed with.

C. For Sorting:
Quick sort:
Despite the fact that the shell sort calculation is altogether superior to anything addition sort, there is still opportunity to get better. A standout amongst the most mainstream sorting calculations is quick sort. Quick sort executes in O (n log n) by and large, and O (n2) in the most pessimistic scenario. Notwithstanding, with legitimate safeguards, most pessimistic scenario conduct is impossible. Quick sort is a non-stable sort.

It is not a set up sort as stack space is required. For further perusing, counsel Carmen. The quick sort calculation works by dividing the cluster to be sorted, then recursively sorting every parcel. In Partition one of the cluster components is chosen as a rotate esteem. Values littler than the rotate esteem are set to one side of the turn, while bigger qualities are set to one side.

IV. CONCLUSION

Proper domain understanding and selection for academic BE projects. The list and description about particular domain gives clear idea to the students for selection process. Aptitude selection will lets students understand their real interests and will provide the resulting domain in which a particular student can at least do research by his/her point of interest.

We also added one more module in which students can share their knowledge, or information related to education with their Classmates as well as teachers.

This system can be used for college level programs. The system deals with time and cost effectiveness of the calculated risks. System works in complex at its back but proves more simple and useful to the students, teachers as well as TPO's (Training and Placement Officers). The resulting solution will be the resulting domain according to test results for individual as well as group of students having common domains after the test. The marks entered

by the Students will be visible to the TPO who will return can send the information as well as mail to the students in the eligibility standards.

FUTURE SCOPE

This type of system can be created for all the 10th or 12th standard students for selecting the future field of education by taking aptitude tests for calculating their knowledge according to particular field. This system will be helpful to know field of interest.

ACKNOWLEDGEMENT

REFERENCES

[1] Efficiency of data mining models to predict academic performance and a co-operativelearning model, IEEE2016
[2] Data mining with decision trees and decision rules, ACM2010
[3] A comparative study for predicting student's academic performance using Bayesian Network Classifiers, IOSRJEN2013
[4] Data mining and analysis, Library of Congress Cataloging in Publication Data, Zaki, Mohammed J.2014
[5] Combining co-operative learning and individual approach in Japanese college, Japanese College EFL Course2014.
[6] Designing a learning model using the STAD technique with a suggestion system to decrease learner's weakness, Elsevier Ltd.2013
[7] web-based learning environment, Elsevier Ltd.2014
[8] Content-free collaborative learning modeling using data mining, Artificial Intelligence Department, E.T.S.I.I., UNED, Ciudad Universitaria, 2013
[9] Personalized Multi-Regression Models for Predicting Students Performance inCourse Activities, University of Minnesota, 2014
[10] Determinants of Students Performance In Intermediate Accounting, Journal of College Teaching and Learning, 2015