Student Performance Analysis Based on Learning

Komal Jirage[1], Rucha Choudhari[2], Varsha Gupta[3], Anil Patthe[4] and Disha Maind[5]

[1]Komal Jirage, Computer Department/ VIIT/ SPPU, Pune, India
[1]jiragekomal82@gmail.com
[2]Rucha Choudhari, Computer Department/ VIIT/ SPPU, Pune, India
[2]choudharirucha@yahoo.in
[3]Varsha Gupta, Computer Department/ VIIT/ SPPU, Pune, India
[3]varshagupta04@yahoo.in
[4]Anil Patthe, Computer Department/ VIIT/ SPPU, Pune, India
[4]pattheanil@gmail.com
[5]Disha Maind, Computer Department/ VIIT/ SPPU, Pune, India
[5]dishamaind@gmail.com

ABSTRACT

Performance analysis of outcome based on learning is a system which will strive for excellence at different levels and diverse dimensions in the field of student's interests. n this paper data clustering technique named k-means clustering is applied to analyze student's learning behavior. The student's evaluation factor like class quizzes, mid and final exam assignment are studied. It is recommended that all these correlated information should be conveyed to the class adviser before the conduction of final exam. This study will help the teachers to reduce the drop out ratio to a significant level and improve the performance of student.

Keywords — Data Mining, Classification, C4.5 Algorithm, Academic Performance.

1. INTRODUCTION

Performance analysis of outcome based on learning is a system which will strive for excellence at different levels and diverse dimensions in the field of student's interests. In this system we are going to find out student trends on the basis of outcomes of tangible records as well as indirect surveys. This data collected will be processed by k-means technique of data mining. A result form k-means algorithm will be recognize as Trend. This trend will help us to track where the students excel and where not and what are their abilities which can be enhanced. The analysis will summarize the outcome and will classify students based on the results. This system will recognize interest of student in particular areas. To empower our student for substantial contribution to economical, technological, entrepreneur and social progress of the society.

.

2. PROBLEM DEFINATION

2.1. Problem Statement

The problem of	The faculty cannot find out students abilities and their interest easily so that they can enhance them in it.
Affects	Thus it may affect with poor university results, placement and career of individual.
The impact of which is	The impact is- it help us from fulfilling mission and vision of the institute.
A successful solution would	If the project get successful then it will be great help for faculty to enhance education system

Table 1. Problem Statement

2.2. Product Position Statement

For	Institute
Who	Faculty
The (product name)	is a (product category)
That	(statement of key benefit - that is - compelling reason to buy)
Unlike	(primary competitive alternative)
Our product	(statement of primary differentiation)

Table 2. Product Position

3. PRODUCT OVERVIEW

3.1. Product Perspective

1. .In presently available system the faculty feel problem at the time of student assessment.
2. Presently there no such system which can find out abilities and skills of students on one particular basis.
3. Faculty has to find out manually the areas of interest of students to excel them in academics as well as extra circulars.
4. We are going to set program outcomes (PO), program educational objective (PEO),mission and vision.
5. System will provide us with a feedback on course recommendations and electives recommendations.
6. There will be two surveys, direct and indirect. Direct will include all the tangible outcomes on student performance and indirect will include student's personal agreements and disagreements.
7. This will result in a summarized solution. This summarized solution will include a result from mapping of direct and indirect surveys and will give us trend.
8. There will be mapping between trend and set objectives of POs
9. There will be mapping between POs and set objectives of POEs
10. The overall results will be compared with mission and vision.

3.2. Project Functions

1. The system is designed to automize the solution needed to our faculty.
2. System enables the direct and indirect surveys of student based on their analytical, designing, coding skill and so on.
3. Our system has designed the form which contains the personal details, educational details, departmental details, knowledge based skills, course recommendation, electives, personal abilities, social awakes, placement, which has to be filled by each students. Our system will have each students academics records .

4. DECISION TREE ALGORITHM

4.1. Data Mining

Data mining [7] is the process of discovering interesting knowledge, such as associations, patterns, changes, significant structures and anomalies, from large amounts of data stored in databases or data warehouses or other information repositories [1]. It has been widely used in recent years due to the availability of huge amounts of data in electronic form, and there is a need for turning such
data into useful information and knowledge for large applications. These applications are found in fields such as Artificial Intelligence, Machine Learning, Market Analysis, Statistics and Database Systems, Business Management and Decision Support[2].

4.1.1. Classification

Classification is a data mining technique that maps data into predefined groups or classes. It is a supervised learning method which requires labeled training data to generate rules for classifying test data into predetermined groups or classes[2]. It is a two-phase process. The first phase is the learning phase, where the training data is analyzed and classification rules are generated. The next phase is the classification, where test data is classified into classes

according to the generated rules. Since classification algorithms require that classes be defined based on data attribute values, we had created an attribute "class" for every student, which can have a value of either "Pass" or "Fail".

4.2. C4.5

C4.5 is a well-known algorithm used to generate a decision trees. It is an extension of the ID3 algorithm used to overcome its disadvantages. The decision trees generated by the C4.5 algorithm can be used for classification, and for this reason, C4.5 is also referred to as a statistical classifier. The C4.5 algorithm made a number of changes to improve ID3 algorithm[2] . Some of these are:

• Handling training data with missing values of attributes
• Handling differing cost attributes
• Pruning the decision tree after its creation
• Handling attributes with discrete and continuous values

Let the training data be a set S = s1, s2 ... of already classified samples. Each sample Si = xl, x2... is a vector where xl, x2... represent attributes or features of the sample. The training data is a vector C = c1, c2..., where c1, c2... represent the class to which each sample belongs to. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits data set of samples S into subsets that can be one class or the other . It is the normalized information gain (difference in entropy) that results from choosing an attribute for splitting the data. The attribute factor with the highest normalized information gain is considered to make the decision. The C4.5 algorithm then continues on the smaller sub-lists having next highest normalized information gain.

5. TECHNOLOGIES USED

5.1. HTML and CSS

HyperText Markup Language (HTML) is a markup language for creating web pages or other information to display in a web browser. HTML allows images and objects to be included and that can be used to create interactive forms. From this, structured documents are created by using structural semantics for text such as headings, links, lists, paragraphs, quotes etc. CSS (Cascading Style Sheets) is designed to enable the separation between document content (in HTML or similar markup languages) and document presentation. This technique is used to improve content accessibility also to provide more flexibility and control in the specification of content and presentation characteristics. This enables multiple pages to share formatting and reduce redundancies.

5.2. MySQL

MySQL is the most popular open source RDBMS which is supported, distributed and developed by Oracle. In the implementation of our web application, we have used it to store user information and student's data[4].

6. IMPLEMENTATION OF C4.5

C4.5 algorithm [5] is a successor of ID3 that uses gain ratio as splitting criterion to partition the data set. The algorithm applies a kind of normalization to information gain using a "split information" value.

a) Measuring Impurity - Splitting Criteria

To determine the best attribute for a particular node in the tree it use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A, relative to a collection of examples S, is defined as

$$\text{Entropy}(S) = -P_j \log_2 P_j$$

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{v \in \text{values}(A)} |S_v|/|S| \ \text{Entropy}(S_v)$$

Where Values (A) is the set of all possible values for attribute A, and Sv is the subset of *S* for which attribute *A* has value *v* (i.e., Sv = {s S | A(s) = v*}*). The first term in the equation for Gain is just the entropy of the original collection *S* and the second term is the expected value of the entropy after S is partitioned using attribute *A*. The expected entropy described by this second term is imply the sum of the entropies of each subset weighted by the fraction of examples | S | / | S | that belong to Gain (S, A) is therefore the expected reduction in entropy caused by knowing the value of attribute A[6].

$$\text{Split Information}(S,A) = -\sum_{i=1}^{n} |S_t|/|S| \ \log_2 \ |S_t|/|S|$$

$$\text{Gain Ratio}(S,A) = \text{Gain Ratio}(S,A) / \text{Split Information}(S,A)$$

7. CURRICULUM DESIGN

The programme curriculum is to be designed such that the students should demonstrate the essential knowledge, skills, and abilities needed for professional practice and higher studies. The curriculum should align with the programme educational objectives through its direct support from programme outcome. The programme curriculum should also satisfy the programme specific criteria. A curriculum design committee is to be formed. The processes may be followed by the committee is as follows.

• **Inputs**

o Program Educational Objectives

o Program Outcomes

o Program specific Criteria

• **Process**

Identify the curricular components that cover depth and breadth for the attainment of programme educational objectives.

The curricular components may include :

_ Humanities and Social Sciences

_ Basic Sciences

_ Engineering sciences

_ Discipline Core

_ Discipline Electives

_ Inter-disciplinary Electives

_ Project

_ Co-curricular and Extra-curricular Activities

Determine the credits for the identified curricular components like Basic Sciences, Humanities &Social Sciences, professional core,electives, projects, co-curricular and extra curricular activities

- Identify the courses/tasks in each curricular component to attain program outcome
- Define the course outcomes for each course and give the correlation with the program outcomes.
- Schedule the course semester-wise and prepare the prerequisite flowchart for the courses in the curriculum
- Obtain the approval of curriculum by competent authorities
- The individual courses would have the following :-
- Department, Course Number and title of Course
- Identification of Course Designers Mapping with Faculty Expertise
- Designation as a Core or Elective course
- Prerequisites
- Contact Hours and type of course (Lecture, tutorial, seminar, project, etc)
- Course Assessment Methods (Both Continuous and Semester-end Assessment
- Course Outcomes
- Topics Covered
- Text Books and/or Reference Material

8. ASSESMENT

Assessment tools are categorized into direct and indirect methods to assess the programme educational objectives, programme outcomes and course outcomes.

Direct methods display the student's knowledge and skills from their performance in the continuous assessment tests, end-semester examinations, presentations, and classroom assignments etc. These methods provide a sampling of what students know and/or can do and provide strong evidence of student learning.

Indirect methods such as surveys and interviews ask the stakeholders to reflect on student's learning. They assess opinions or thoughts about the graduate's knowledge or skills. Indirect measures can provide information about graduate's perception of their learning and how this learning is valued by different stakeholders.

1. Alumni survey(Indirect)

Collection of a wide variety of information about program satisfaction, how well students are prepared for their careers, what types of jobs or graduate degrees majors have gone on to obtain, and the skills that are needed to succeed in the job market or in graduate study, 3 years after the graduation. Provide the information opportunity to collect data on which areas of the program should be changed, altered, improved or expanded

A. Student Exit survey (Indirect)

To evaluate the success of the programme in providing students with opportunities to achieve the programme outcomes.

B. Project Evaluation (Direct)

This is a demonstration of the abilities of a student throughout the programme.

2. Assessment of PEOs:

Program Educational Objectives:

Graduates of our program will

1. Provide students with sound knowledge of Mathematics, Science and Technology to build a logical base of Computer engineering that will be useful in solving complex engineering problems and develop enduring learning ability.

2. Impart knowledge with good understanding of fundamentals of all subjects of Computer Engineering, so that students are able to analyze, design and implement new projects from various application domains using modern engineering tools.

3. Develop logical thinking & programming skills to enable students to design, develop system and application level software within realistic constraints.

4. Inculcate among students ethics, social responsibilities and professionalism.

5. Improve communication, presentation, entrepreneurial skills and team work leading to competent professionals and address challenges of IT scenarios at global level.

Mission Attributes	PEO 1	PEO 2	PEO 3	PEO 4	PEO5
Professionalism			**X**	**X**	**X**
Global issues					**X**
Technical Skills	**X**	**X**	**X**		
Higher Studies	**X**				
Industry needs			**X**		**X**
Research		**X**	**X**		
Entrepreneurship					**X**
Ethics and social responsibilities				**X**	

Programme Outcomes

(a) Apply knowledge of mathematics, science, and engineering to solve the problems in design, modelling and analysis in computer based systems.

(b) Apply the design principles, conduct experiments, analyze and interpret data for software and hardware systems by means of various mini projects.

(c)Design a system, process to meet desired needs within realistic constraints such as economic, social, health and sustainability by means of curricular and extracurricular activities.

(d) Work as member of project team to find successful solutions in the area of soft computing and embedded systems.

(e) Design and solve real time problems and data analysis in domain specific problems.

(f) Understand the professional efficiency and ethical responsibility.

(g) Communicate effectively in engineering community at large by means of effective presentations, report writing, paper publications.

(h) Understand the impact of engineering solutions in a global economic and societal context.

(i) Recognize the opportunities required to engage in lifelong learning and overall development as demonstrated through.

PEOs	a	b	c	d	e	f	g	h	i	J	k
1	x				x				x		
2		x	x	x				x		X	x
3	x		x		x						

A. Mapping of program outcomes with academic attainments

B. Mapping of program outcomes with professional attainments

C. Mapping of program outcomes with academic and professional contributions by faculty

D. Mapping of program outcomes with placement

9. CONCLUSIONS

In this paper, an approach for efficient retrieval of clustered search results has been proposed, in which the similarity between the documents can be compared by considering the co-occurrence term of a query and not just by the contents of a document. All the documents are compared and the resultant clusters are formed by using K-Means clustering algorithm which improves the relevancy rate and processing time. In this study we make use of data mining process in a student's database using k-means clustering algorithm to predict student's learning activities. We hope that the information generated after the implementation of data mining technique may be helpful for instructor as well as for students. This work may improve student's performance; reduce failing ratio by taking appropriate steps at right time to improve the quality of education. For future work, we hope to refine our technique in order to get more valuable and accurate outputs, useful for instructors to improve the students learning outcomes.

ACKNOWLEDGMENT

Author take this opportunity to express our deepest sense of gratitude and sincere thanks to those who have helped us in completing this task

REFERENCES

[1] Han, J. and Kamber, M., (2006) Data Mining: Concepts and Techniques, Elsevier.

[2] Dunham, M.H., (2003) Data Mining: Introductory and Advanced Topics, Pearson Education Inc.

[3] Kantardzic, M., (2011) Data Mining: Concepts, Models, Methods and Algorithms, Wiley-IEEE press.

[4] MySQL – The world's most popular open source database, http://www.mysql.com/

[5] J.R.Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, Inc, 1992.

[6] Quinlan, "Learning decision tree classifiers". ACM Computing Surveys (CSUR), 28(1):71–72, 1996.

[7] J.R.Quinlan, "Induction of Decision Tree", Journal of Machine learning", Morgan Kaufmann Vol.1, 1986, pp.81-106.

[8] J. Quinlan, "Learning decision tree classifiers". ACM Computing Surveys (CSUR), 28(1):71–72, 1996.

[9] Shreenath Acharya, Madhu N, "Discovery of students' academic patterns using data mining techniques" IJCSE,Vol. 4 No. 06 June 2012.

