Using Computers as a Learning Tool by the Students of Arts Disciplines: A Data Mining Case Study

V. Selvakani*, V. Annalakshmi, H. Archana
PG Student, Asst. Professor, PG Student
Department of Computer Science, JayarajAnnapackiam College for Women (Autonomous),
Periyakulam, Theni Dt, Tamil Nadu, India

Abstract— The rapid development in computer technology and the wide availability of personal computers together with the Internet, email, and various literature retrieval applications have changed both, the study and the practice environments in all disciplines. Computers can support the variety of ways learners construct their own understanding. Students who gather information from the Internet can be self-directed and independent. Computer-assisted learning (CAL) in education is rapidly increasing. Now-a-day, there is general awareness of the potential benefit of CAL all over the world, including the developing countries, and many faculties recognize the need to exploit the capacities of information communication technology (ICT) to enhance their educational programs. Observations on student's competence with ICT are few, and are carried out mostly in countries where informatics is well developed. Hence, this study was undertaken to investigate the current knowledge and skills of undergraduate arts students of Arts and Science colleges of Theni district, Tamil Nadu in India with respect to Information Communication Technology (ICT). We have collected the information from many students who are using computers for their academic purposes. This process is done by using the classification algorithm in data mining and we select with the XL Miner tool.

Keywords— Computers; Learning Tool; College Students; Data Mining; Theni District

I. INTRODUCTION

The most popular buzz word in the modern era is "Technology". Today without technology, it is difficult to lead one's life peacefully. It played a vital role in even teaching and learning techniques. Technology also has the power to transform teaching by ushering in a new model of connected teaching. It infuses classrooms with digital learning tools, such as computers and hand held devices; expands course offerings, experiences, and learning materials; supports learning 24 hours a day, 7 days a week; builds 21st century skills; increases student engagement and motivation and accelerates learning[1]. This model links teachers to their students and to professional content, resources, and systems to help them improve their own instruction and personalize learning [2]. We intended to conduct a survey to investigate the current knowledge of computers among the undergraduate students of arts disciplines in Arts and Science colleges of Theni district. The purpose of the survey was to gather information not only about computer skills, but also about overall comfort level in using computers and about expectations for enhancing computer skills while attending computer courses at college level. The main objective of the case study is to find out the usages of the computers for learning purposes by the Arts disciplines students. The data was collected from many Arts and Science colleges of Theni district and analyzed using XL-Miner tool of the Data mining. We identified that the usages for academic purposes by the rural college arts disciplines students are very less even though they have government issued free laptops are with them. But it was found that they gradually improving their skills in operating computers by the help of these free laptops.

II. PROBLEM DEFINITION AND DESCRIPTION

Theni, one of the southern districts of Tamilnadu state, is bounded on the north by Dindigul district, on the east by Madurai district, on the south by portions of Virudhunagar district and Idukki district of Kerala State and on the west by Idukki (Kerala). The total geographical area of the district is 3076.30 Sq. Km. According to the 2011 census, it has a population of 1,243,684. The district has a population density of 433 inhabitants per square kilometre (1,120 /sq. mi.). Its population growth rate over the decade 2001-2011 was 13.69 % [3]. The literacy level of Theni district according to figures available for the year 2011 is 77.62 %. It is also observed that the male literacy level has grown steadily from 57.40 % in 1981 to 85.48 % in 2011. There has also been a significant increase of female literacy level from 31.68 % in 1981 to 69.72 % in 2011.

It has only one Medical College, six Engineering Colleges, eleven Arts and Science Colleges, two Nursing Colleges, seven Colleges of Education, Horticulture College and College of Physical Education are each one. At most 18640 students are doing their graduation in this district of which 8485 are male students and 10155 are female.

Table 1: Institution and Student Strength Details as of April, 2014

S.No.	Institutions Details	No. of			Total Strength
		Institution(s)	Boys	Girls	
1.	Medical College	1	125	175	300
2.	Engineering Education	6	3570	1820	5390
3.	Arts & Science College	11	4500	7460	11960
4.	Horticulture College	1	130	120	250
5.	Nursing Colleges	2	-	100	100
6.	College of Physical Education	1	-	80	80
7.	Colleges of Education	7	160	400	560
	Total	**29**	**8485**	**10155**	**18640**

Source: Concern College Principals

III.   METHODOLOGY

Methods play a major role in research. This study has adopted the survey method and descriptive research design. The population of the study consists of the entire set of student population in the Arts and Science colleges of Theni district. According to the statistical data from Theni chief educational office, a total of 11960 students are studying in eleven Arts and Science colleges of Theni district of which 4500 are male and 7400 are female students. A total of 3857 respondents of the survey region have been covered as the sample. Data were gathered using a self-administered questionnaire prepared specially for these purpose.

The following interview questions brought a wealth of information to the researchers.

- Do you currently have regular access to a computer?
- Do you have an e-mail account?
- Do you use keyboard shortcuts?
- Can you properly shut down a computer?
- Can you perform file management including deleting and renaming files, etc.?
- Can you use a 'search' command to locate a file?
- Can you write files onto a CD or USB?
- Can you print a document using a printer?
- Can you copy, cut and paste text in a document?
- Can you create a basic Word document?
- Can you create a basic Excel spreadsheet?
- Can you create a simple presentation using PowerPoint?
- Can you download and save files from the Web?
- I enjoy using computers.
- I think that computers are difficult to use.

IV.   DATA MINING PROCESS

Data mining is defined as a process of discovering hidden valuable knowledge by analyzing large amounts of data, which is stored in databases or data warehouse, using various data mining techniques such as machine learning, artificial intelligence (AI) and statistical. Therefore the needs for a standard data mining process increased dramatically.

Any data that was collected by any means must be pre-processed to bring it to a form suitable for pattern recognition. Starting with the raw data in the form of images or meshes, we successively process these data into more refined forms, enabling further processing of the data and the extraction of relevant information. Detailed transaction information in the OLTP (On Line Transaction Processing) and legacy system is usually not suitable for data mining. For data mining to be effective, much careful work is needed in defining the aims of network of data mining then in selection, cleaning, transformation and separate storage of data that is suitable for data mining. A typical data mining process includes requirement analysis, data selection and collection, cleaning mining exploration and validation, implementing, evaluating and validation, monitoring and result visualization[4]. We implement classification algorithm to evaluate the predicted data of our case study. We generated the reports for result visualization and exploration.

4.1 Requirement Analysis: There are some goals that the data mining process is expected to achieve. The samples of the case study must be clearly defined. In requirement analysis, the problem is clearly defined as the objective of the case study. Objective of this case study is to gather information not only about computer skills, but also about overall comfort level in using computers and about expectations for enhancing computer skills while attending computer courses at college level in the Arts disciplines students Arts and Science Colleges of Theni district.

4.2 Data Selection and Collection: This step includes finding the best source of databases for the analysis. If the data has implemented a data warehouse, then most of the data could be available there. If the data is not available in the warehouse, the source OLTP (On Line Transaction Processing) systems need to be identified and the required information extracted and stored in some temporary system. In some cases, only a sample of the data available may be

required [5]. In our case study, data were collected using questionnaire from the study area. Questionnaires were used to survey the reason the overall comfort level in using computers and about expectations for enhancing computer skills of the Arts disciplines students Arts and Science Colleges of Theni district.

4.3 Cleaning and Preparing Data: This may not be an onerous task if a data warehouse contains the required data, since most of this must have already been done when data was loaded into the warehouse. Otherwise this task can be very resource intensive and sometimes more than 50% of effort in a data mining project is spent on this step. Essentially, a data store that integrates data from a number of databases may need to be created. When integrating data, one often encounters problems like identifying data, dealing with missing data, data conflicts and ambiguity. An ETL (Extraction, Transformation and loading) tool may be used to overcome these problems.

4.4 Data Mining Exploration and Validation: Once appropriate data has been collected and cleaned, it is possible to start data mining exploration. Assuming that, the user has access to one or more data mining tools, a data mining model may be constructed based on the sample details. It may be possible to take a sample of data and apply a number of relevant techniques. For each technique, the results should be evaluated and their significance is interpreted. This is likely to be an iterative process which should lead the selection of one or more techniques that are suitable for further exploration, testing and validation.

4.5 Implementing, Evaluating and Monitoring: Once a model selected and validated, the model can be implemented by the decision makers. This may involves software development for generating reports or for results visualization and exploration which is used by the managers. It may be more than one technique that is available for the given data mining task. It is important to evaluate the results and choose the best technique. Furthermore, there is a need for regular monitoring the performance of the techniques that have been implemented [6].

In our case study, we have implemented Classification Algorithm to evaluate the predicted data. We generated the reports for result visualization and exploration. It may be more than one technique is available for given data mining task.

4.6 Result Visualization: Here, we classify the student's computer skills according to the category. Although the problem is dealing with number of dimensions must visualized using two-dimensional computer screen or printout. And also see that the decision tree making of that problem. We derived the decision tree out of our processing and we have given it in Fig. 1.
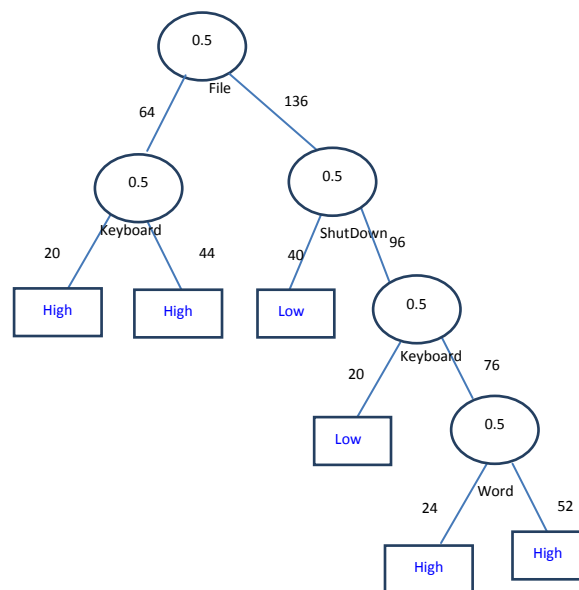

Fig. 1. Classification Full Tree

V. DECISION TREE ALGORITHM

Algorithm: Generate decision tree. Generate a decision tree from the training tuples of data partition D.

Input: Data partition, D, which is a set of training tuples and their associated class labels; attribute list, the set of candidate attributes; Attribute selection method, a procedure to determine the splitting criterion that "best" partitions the data tuples into individual classes. This criterion consists of a splitting attribute and, possibly, either a split point or splitting subset [7].

Output: A decision tree.

Method:

        Step 1:   Create a node N;

        Step 2:   If tuples in D are all of the same class, C then

        Step 3:   Return N as a leaf node labeled with the class C;

Step 4:   If attribute list is empty then
Step 5:   Return N as a leaf node labeled with the majority class in D;

// majority voting
Step 6:   Apply Attribute selection method(D, attribute list) to find the "best" splitting criterion;
Step 7:   Label node N with splitting criterion;
Step 8:   If splitting attribute is discrete-valued and multi way splits allowed then

// not restricted to binary trees
Step 9:   Attribute list attribute list □ splitting attribute; // remove splitting attribute
Step 10:   For each outcome j of splitting criterion

// partition the tuples and grow sub trees for each partition
Step 11:   Let Dj be the set of data tuples in D satisfying outcome j; // a partition
Step 12:   If Dj is empty then
Step 13:   Attach a leaf labelled with the majority class in D to node N;
Step 14:   Else attach the node returned by Generate decision tree(Dj, attribute list)
                to node N; endfor
Step 15:   Return N;

VI.   ALGORITHM IMPLEMENTATION

In this case study, we have classified the student's computer skills according the category we derived using data mining techniques. This study has adopted the survey method and descriptive research design. The population of the study consists of the entire set of student population in the Arts and Science colleges of Theni district. According to the statistical data from Theni chief educational office, a total of 11960 students are studying in eleven Arts and Science colleges of Theni district of which 4500 are male and 7400 are female students. A total of 3857 respondents of the survey region have been covered as the sample.The information gathered by means of questionnaire was also verified by personal interviews. The result is then compiled for both quantitative as well as qualitative data.

Step 1: Input variables are identified. Here we declare Var1 as keyboard,Var2 as shutdown, Var3 as file, var3 as word, Var4 as cu/co/pa. Then the output we have taken is the tuple yes as 1, No as 0.

Input					
Data					
Training data used for building the model	['computer.xlsx']'Sheet1'!$A$2:$S$201				
# Records in the training data	3857				
Input variables normalized	No				
Variables					
# Input Variables	5				
Input variables	Keyboard	ShutDown	File	Word	Cu/Co/Pa
Output variable	Class				
Parameters/Options					
Early stopping of tree growth required		Yes			
Minimum # records in a terminal node		20			
Is pruning done		No			
Max # levels displayed in tree drawing		5			
Draw full tree		Yes			

Fig. 2. Input data and variables

Step 2: Here, we find out the probability value for each attribute by using information gained. The information gain is calculated from the Following Formula,

$$\text{Info (D)} = -\{n_{i=1}P_i \log_2(P_i)$$

Where Pi is the Probability that an arbitrary tuple belongs to the class. Here all the classes are in the probability value of 0.3.

Prior class probabilities	
Class	**Prob.**
High	0.333333333
Low	0.333333333
Middle	0.333333333

Fig. 3. Prior class probabilities

Step 3: Training log is used to find out the miss- classify of the given class. It also used for growing the full tree using data.

Training Log (Growing the full tree using training data)	
# Decision Nodes	% Error
0	44
1	44
2	44
3	42
4	40
5	40

Fig. 4. Training Log

Step 4: The full tree rules are used for find out the decision and terminal nodes, where the decision nodes are 5 and terminal nodes are 6.

Full Tree Rules (Using Training Data)									
#Decision Nodes		5			#Terminal Nodes		6		
Level	NodeID	ParentID	SplitVar	SplitValue	Cases	LeftChild	RightChild	Class	Node Type
0	0	N/A	File	0.5	200	1	2	High	Decision
1	1	0	Keyboard	0.5	64	3	4	High	Decision
1	2	0	ShutDown	0.5	136	5	6	High	Decision
2	3	1	N/A	N/A	20	N/A	N/A	High	Terminal
2	4	1	N/A	N/A	44	N/A	N/A	High	Terminal
2	5	2	N/A	N/A	40	N/A	N/A	Low	Terminal
2	6	2	Keyboard	0.5	96	7	8	High	Decision
3	7	6	N/A	N/A	20	N/A	N/A	Low	Terminal
3	8	6	Word	0.5	76	9	10	High	Decision
4	9	8	N/A	N/A	24	N/A	N/A	High	Terminal
4	10	8	N/A	N/A	52	N/A	N/A	High	Terminal

Fig. 5 (a). Tree rules using training data

Training Data scoring - Summary Report (Using Full Tree)			
Classification Confusion Matrix			
	Predicted Class		
Actual Class	High	Low	Middle
High	96	16	0
Low	16	24	0
Middle	28	20	0
Error Report			
Class	# Cases	# Errors	% Error
High	112	16	14.29
Low	40	16	40.00
Middle	48	48	100.00
Overall	200	80	40.00

Fig. 5 (b). Error report

Step 5: The tree was constructed with corresponding attributes. The tree has taken the root node as the attribute Economic status. The root node is selected by the probability value. Finally in this step, to classify the actual class to predicted classes in given data such as keyboard, shutdown, file and word will get from this step. The overall elapsed time to run the XL Miner for this case study is 6 Sec.
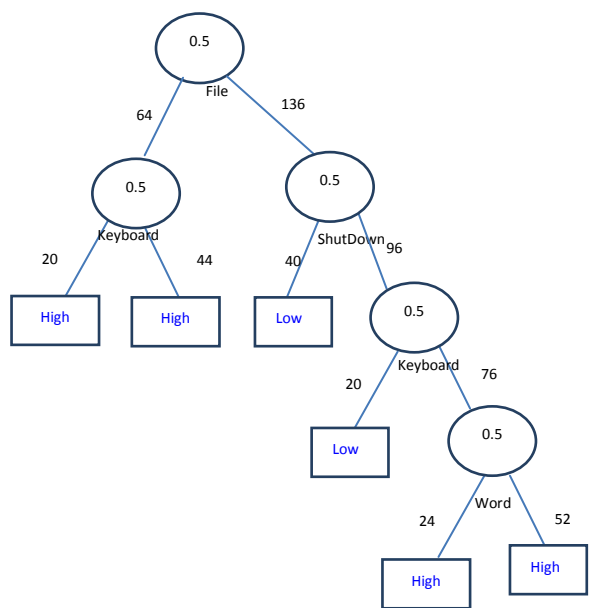
Fig. 6. Classification full tree using training data

VII.   DATA MINING FINDINGS

The initial studies unveiled a number of relationships between variables as well as threshold values that justify further analysis. Several attributes are useful predictor of retention and attritions. These explanations increase confidence that the values of the attributes will continue to be predictors in future.

VIII.   CONCLUSION

The goal of data mining classification is to build a set of models that can correctly predict the class of the different objects. The input to this method is set of objects (training data); the classes in which these objects belong to (dependent variables), and a set of variables describing different characteristics of the objects (independent variables). Before the case study was done, it is very difficult to evaluate the comfort level of using computers by the arts disciplines students. After completing this case study, it is very useful to classify the reasons for using the computer [8]. We found out that the comfort level of using computers for academic purposes by the arts disciplines students of Arts and Science colleges of Theni district is high. The basic knowledge that are necessary to use the computers are developed by the free laptops given by government of Tamilnadu and the usages of computers by the ordinary students become high in Theni district.

REFERENCES

[1]   "*Use of Technology in Teaching and Learning*", available at:http://www.ed.gov/oii-news/use-technology-teaching-and-learning.

[2]   Margaret Honey,Katherine McMillan Culp and Robert Spielvogel, "*Critical Issue: Using Technology to Improve Student Achievement*", North Central Regional Educational Laboratory, 2005

[3]   'District Census 2011'. Census2011.co.in. 2011, www.census2011.co.in/district.php.

[4]   Gupta G. K.,2011, "Introduction to Data Mining with Case Studies", 2ndEdision, PHI Learning publication, Page 7.

[5]   AgarwalR,ImielinskiT,MannilaH,ToivonenHVerkamoAI,"Data mining classification Rules", VLDB, Sept 12-1-1999.

[6]   Isabelle Guyon, "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research 3, 2003.

[7]   Yael Ben-Haim and Elad Tom-Tov, "*A Streaming Parallel Decision Tree Algorithm",* Journal of Machine Learning Research 11 (2010) 849-872 .

[8]   BalakrishnanMuniandy, "*Academic Use of Internet among Undergraduate Students: APreliminary Case Study in a Malaysian University*", International Journal of Cyber Society and Education, Pages 171-178, Vol. 3, No. 2, December 2010.