A Framework for Recommendation of courses in E-learning System

Sunita B Aher
M.E. (CSE) -II
Walchand Institute of Technology
Solapur University India

Lobo L.M.R.J.
Associate Professor, Head, Department of IT
Walchand Institute of Technology
Solapur University India

ABSTRACT
The course recommendation system in e-learning is a system that suggests the best combination of subjects in which the students are interested.

In this paper, we propose a framework for recommendation of courses in the E-learning system. In our approach we collect the data for example student enrollment for a specific set of course. After getting data, we use different combination of algorithm & we analyze the suitability of combination applied for recommendation. In this paper we outline our architecture & we apply the association rule mining at preliminary stage.

Key words:
Moodle, Weka, classification, association rule, clustering algorithm.

1. INTRODUCTION
Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [9].E-learning is the computer and network-enabled transfer of knowledge & skills.

The basic idea is to find the best combination of subjects which will lead to more effective learning in a particular stream. It will also help to understand the behavior of the student that is in which subjects, students are more interested. This result can also be used for designing the new interdependent courses.

The rest of the document is organized as follows. Section 1.1 provides a background of the related research fields covering a brief introduction about each. Section 2 describes the Literature Review. Section 3 discusses the architecture of proposed system. Section 4 discusses result & implementation. The Discussion & future work is presented in Section 5 followed by references.

1.1 Background
This research integrates issues from the research field of Moodle, WEKA and Data mining (Classification, clustering & Association Algorithm).The following subsections includes a brief overview of these topics.

1.1.1. Moodle (Modular Object-Oriented Developmental Learning Environment)
Moodle is an open-source course management learning system to help educators create effective online learning communities [2]. It is also possible to modify the source code of any file of Moodle. It is very easy to add the course in the system. Here are adding 13 course category & near about 82 subjects which are related to Computer Science & Engineering and Information Technology Department. Here we are using the Moodle as it is very user friendly. Also it maintains detailed logs of all activities of students. It stores the record of every click that students make. We can use these logs to find courses in which student are interested. It stores the logs in relational database MYSQL.

Moodle can be used in many types of environment such as in education, training and development, and business settings. Moodle is shown in figure 1.

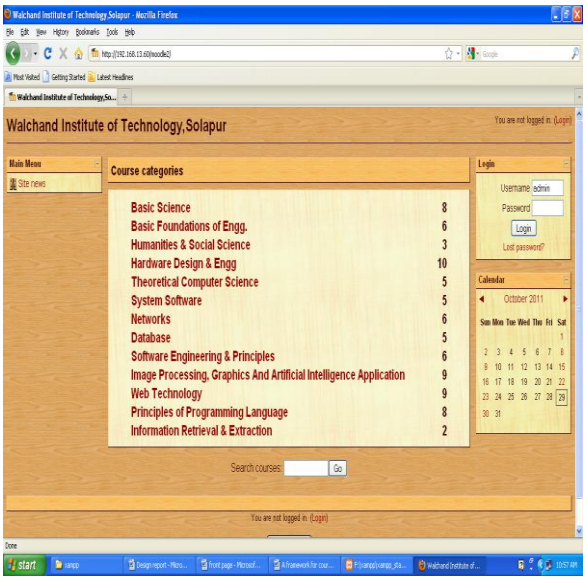

Figure 1: Moodle

1.1.2 Weka
Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code [8]. The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality [10]. It is freely available software. It is portable & platform independent because it is fully implemented in the Java programming language and thus runs on almost any platform. Weka has several standard data mining tasks, data preprocessing, clustering, classification, association, visualization, and feature selection. The WEKA GUI chooser launches the WEKA's graphical environment which has six buttons: Simple CLI, Explorer, Experimenter, Knowledge Flow, ARFFViewer, & Log. Figure 2 shows Weka 3.5.3 with Explorer window.

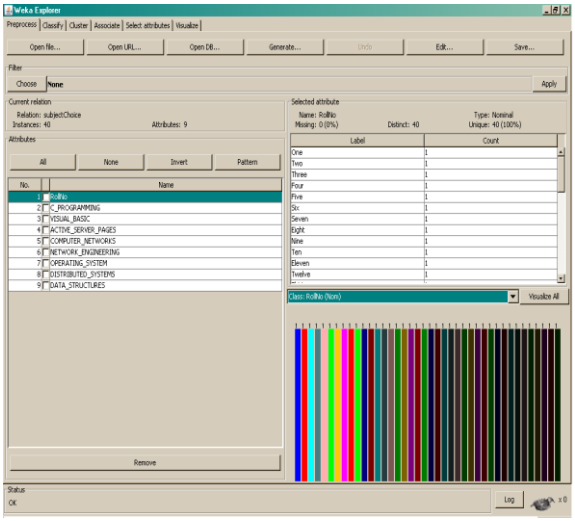

Figure 2: Weka 3.5.3 with Explorer window

1.1.3 Data Mining

Data mining is the process of discovering interesting knowledge from large amount of data stored in database, data warehouse or other information repositories. It includes various tasks such as classification, clustering, association rule etc.

1.1.3.1 Association Rule

Association rules are used to show the relationship between data items. Mining association rules allows finding rules of the form: *If antecedent then (likely) consequent* where *antecedent a*nd *consequent* are itemsets which are sets of one or more items. Association rule generation consists of two separate steps: First, minimum support is applied to find all frequent itemsets in a database. Second, these frequent itemsets and the minimum confidence constraint are used to form rules[10].

Support & confidence are the normal method used to measure the quality of association rule. Support for the association rule X->Y is the percentage of transaction in the database that contains XUY [7]. Confidence for the association rule is X->Y is the ratio of the number of transaction that contains XUY to the number of transaction that contain X [7]. Association rule can be used in educational data mining for analyzing the learning data.

1.1.3.2 Classification

Classification is a data mining task that maps the data into predefined groups & classes. It is also called as supervised learning .It consists of two steps:

1. Model construction: It consists of set of predetermined classes. Each tuple /sample is assumed to belong to a predefined class. The set of tuple used for model construction is training set. The model is represented as classification rules, decision trees, or mathematical formulae. This model is shown in figure 3.
2. Model usage: This model is used for classifying future or unknown objects. The known label of test sample is compared with the classified result from the model. Accuracy rate is the percentage of test set samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur [10].

1.1.3.3 Clustering

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group. Clustering can be considered the most important unsupervised learning technique. In educational data mining, clustering has been used to group the students according to their behavior e.g. clustering can be used to distinguish active student from non-active student according to their performance in activities [10].

2. LITERATURE REVIEW

Castro [1] provides an up-to-date snapshot of the current state of research and applications of Data mining methods in e-learning. It provides the taxonomy of e-learning problems to which the Data Mining techniques have been applied including, for instance: Students' classification based on their learning performance; detection of irregular learning behaviors; e-learning system navigation and interaction optimization; clustering according to similar e-learning system usage; and systems' adaptability to students' requirements and capacities. This paper presents the research along the axis of the data mining modeling techniques & methods e.g. the classification problem in e-learning, association rule, the clustering problem in e-learning, prediction techniques, visualization techniques etc. This paper also present the surveyed content along the e-learning application axis e.g. research work that perform students learning assessment, research work that offer course adaptation based on student's learning style, data mining application providing an evaluation of learning material, data mining application providing feedback to e-learning actors, data mining application for the detection of atypical learning behavior etc. In this paper they presents some tables which are very useful as e-learning material e.g. E-learning projects in which Data Mining techniques are used, e-Learning discussion forums, e-Learning organizations, Key e-learning research papers, Key e-learning books and books chapters, e-Learning information repositories, Learning objects repositories, e-Learning standards, Open source e-learning software etc.

C. Carmona [3], they propose proposed how to use adaptive machine learning algorithms to learn about the student's preferences over time. First they use all the background knowledge available about a particular student to build an initial decision model based on learning styles. This model can then be fine-tuned with the data generated by the student's interactions with the system in order to reflect more accurately his/her current preferences.

In paper [4], Data Mining System (DMS) was designed and implemented to analyze the study records of two programming courses in a distance curriculum of Computer Science. Various data mining schemes, including the linear regression and probabilistic models, were applied to describe and predict student performance. The results mentioned in this paper indicate that a DMS can help a distance education teacher, even in courses with relatively few students, to intervene in a learning process at several levels: improving exercises, scheduling the course, and identifying potential dropouts at an early phase.

Zaı¨ane, O [5] suggests the use of web mining techniques to build recommender system that could recommend on-line learning activities or shortcuts in a course web site based on learners' access history to improve course material navigation as well as assist the online learning process. These techniques are considered integrated web mining as opposed to off-line web mining used by expert users to discover on-line access patterns. They have proposed an approach to build a software agent that uses data mining techniques such as association rules mining in order to build a model that represents on-line

user behaviors, and uses this model to suggest activities or shortcuts. They tested this recommender system approach on an on-line course and will evaluate the recommendations using questionnaires as well as a log that is keeping track of selected recommendations by the users. The approach is also tested on an on-line system used by novice health care providers at the university hospital at the University of Alberta and will be evaluated based on the time-saving recorded for users who follow the suggested shortcuts in comparison with those that ignore the recommendations.

Resende [6] discusses the use of Data Warehouse and Data Mining resources to aid in the assessment of distance learning of students enrolled in distance courses. Information considered relevant for the assessment of distance learning is presented, as is the modeling of a data warehouse to store this information and the MultiStar environment, which allows for knowledge discovery to be performed in the data warehouse.

In paper [11], they compare different data mining methods and techniques for classifying students based on their Moodle usage data and the final marks obtained in their respective courses. They have developed a specific mining tool for making the configuration and execution of data mining techniques easier for instructors.

3. ARCHTECTURE OF PROPOSED SYSTEM

Here in this paper, we are going to consider the 13 course category which is shown in following table 1.

Table 1: Course category & cources for proposed System

Sr No	Course Category	Course Name & Code
1	Basic Science (BS)	1.Engineering Physics(PHY) 2.Engineering Mathematics-I (M-I) 3.Engineering Mathematics-II (M-II) 4.Applied Mathematics-I (AM-I) 5.Applied Mathematics-II (AM-II) 6.Engineering Chemistry (CHEM)) 7.Discrete Mathematical Structure (DMS) 8.Applied Mechanics (MECH)
2	Basic Foundations of Engineering (BFE)	1.Engineering Graphics(EG) 2.Basic Electronics And Computer Programming (BECP) 3.Workshop Practice (WP) 4.Basic Civil Engineering (BCE) 5.Basic Electrical Engineering (BEE) 6.Basic Mechanical Engineering (BME)
3	Humanities & Social Science (HSS)	1.Communication Skills (COMM) 2.Environmental Sciences (ES) 3.Management Information System (MIS)
4	Hardware Design & Engineering (HDE)	1.Switching Theory And Logic Design (STLD) 2.Microprocessor (MP) 3.Advanced Microprocessor (AMP) 4.Computer Organization (CO) 5.Advanced Computer Architecture (ACA) 6.VLSI Technology (VLSI) 7.Microcontroller (MICRO) 8.Digital Signal Processing (DSP)
		9.Mobile Computing (MC) 10.Natural Language Processing (NLP)
5	Theoretical Computer Science (TCS)	1.Data Structures –I (DS-I) 2.Data Structures –II (DS-II) 3.Design & Analysis of Algorithm (DAA) 4.Formal System And Automata (FSA) 5. Theory Of Computation (TOC)
6	System Software (SS)	1.System Programming (SP) 2.Compiler Construction (CC) 3.Operating System – I (OS-I) 4.Operating System – II (UNIX) (OS-II) 5.Distributed Systems (AOS)
7	Networks (NT)	1.Computer Networks – I (CN-I) 2.Computer Networks – II (CN-II) 3.Information Security (IS) 4.Network Engineering (NE) 5.Network Security (NS) 6.Computer Network Administration (CNA)
8	Database (DB)	1.Database Engineering (DBE) 2.Advanced Database System (ADS) 3.Advanced Database Design (ADD) 4.Oracle (SQL) 5.Bioinformatics(BIO)
9	Software Engineering & Principles (SEP)	1.Software Engineering (SE) 2.Object Oriented Modeling And Design (OOMD) 3.Software Testing & Quality Assurance (STQA) 4.Information Technology (IT) 5.Principles Of Management & Engineering Economics (PMEE) 6. Component Technology (CT)
10	Image Processing, Graphics And Artificial Intelligence Application (IP&AIP)	1.Computer Graphics (CG) 2.Image Processing (IP) 3.Embedded System (ESYS) 4.Pattern Recognition (PR) 5.Information Retrieval (IR) 6.Artificial Intelligence (AI) 7.Artificial Neural Network (ANN) 8.Human Computer Interfaces (HCI) 9.Expert System (ES)
11	Web Technology (WT)	1.Hyper Text Markup Language (HTML) 2.Dynamic Html (DHTML) 3.JavaScript (JS) 4.Extensible Markup Language (XML) 5.Java Server Pages (JSP) 6.Servlets (STS) 7.Active Server Pages (ASP) 8.ASP.NET (ASP.NET) 9.Hypertext Preprocessor (PHP)
12	Principles of Programming Language(PPL)	1.Visual Basic (VB) 2.C-Programming (CP) 3.Advanced C (AC) 4.Object Oriented Design & Programming (C++) (OODP) 5.Java Programming (JP) 6.Advanced Java (AJ) 7.Visual C++ (VC++) 8.C-Sharp (CS)

13	Information Retrieval & Extraction (IRE)	1.Data Warehousing (DW) 2.Data Mining (DM)

Under each category there will courses. So there are about 82 subjects. We will create the student login & give the access to the student. Student will enroll for those subjects in which they are interested. This data is stored in the moodle database which we use to find out the best combination. The framework for this recommendation system is shown in the figure 3.

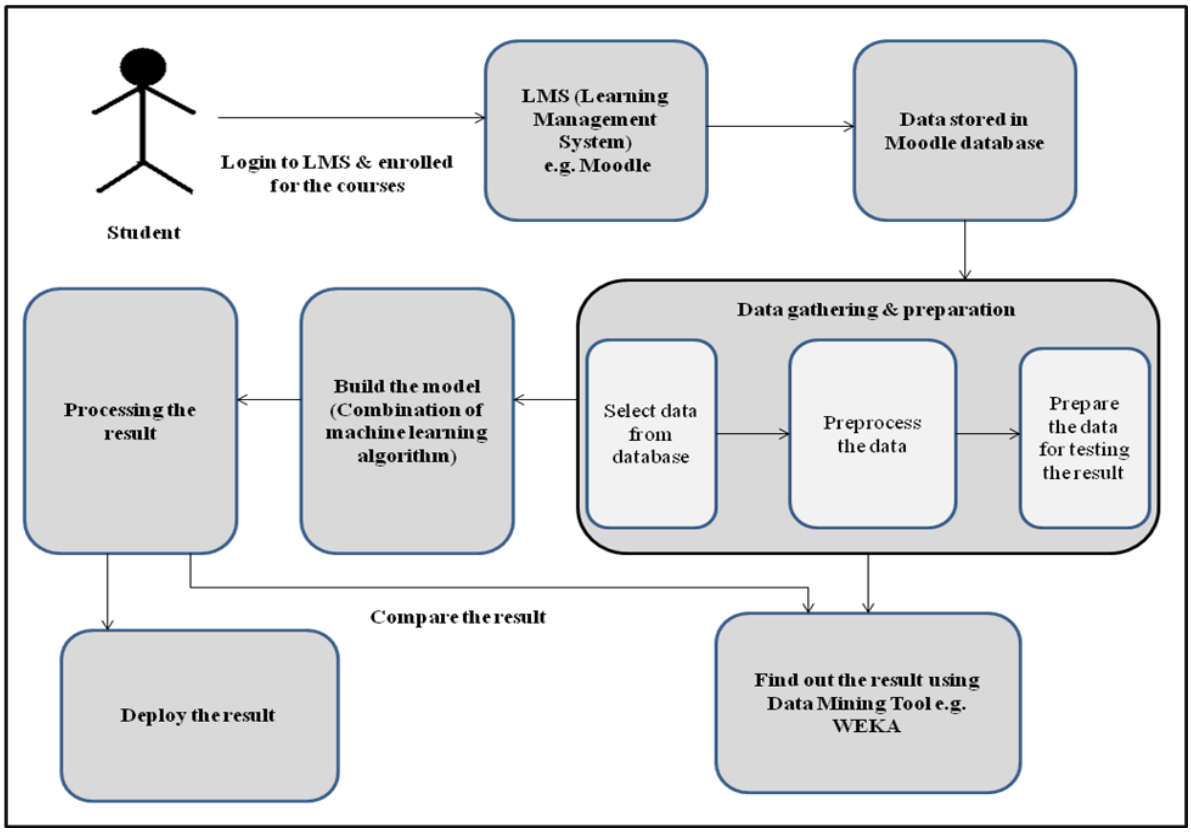

Figure 3: Architecture of Recommendation System

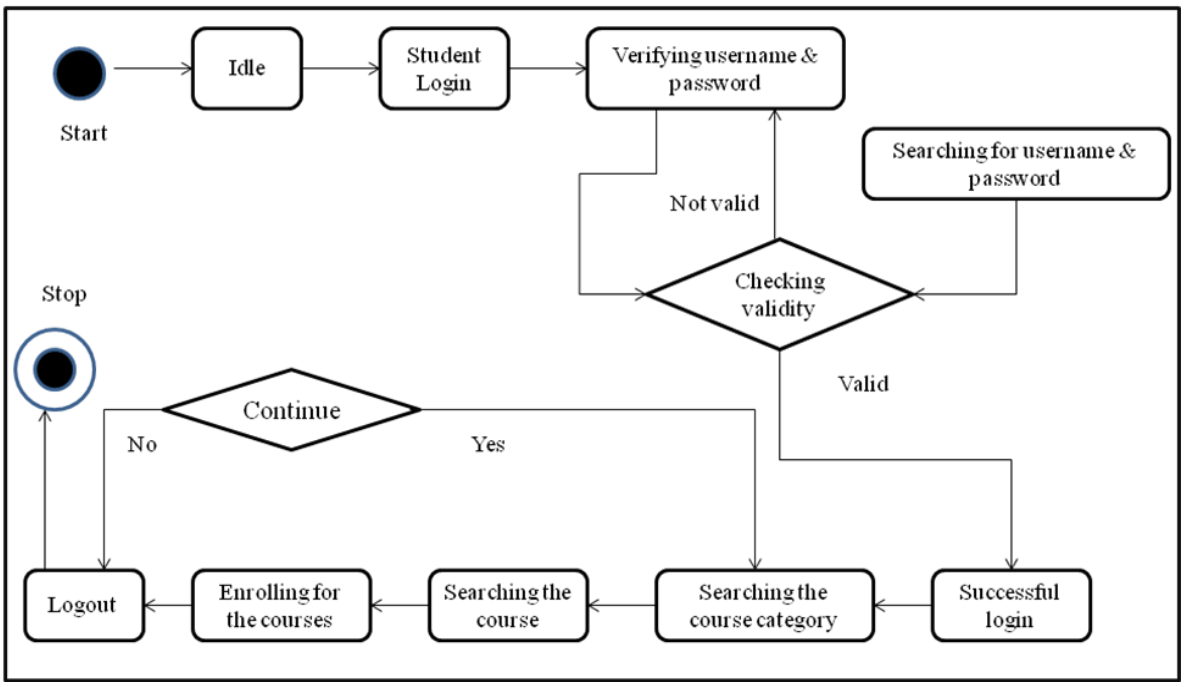

Figure 4: State Machine Diagram for Learning Management System (Moodle)

In this architecture ,student first logs in the Learning Management System e.g. Moodle. The system verifies the username & password. After verifying the username & password, student will search the course category & courses (subjects). Students will enroll for subject in which they are interested. This enrollment information is stored in database as Moodle store each & every action of student. The state machine diagram for Learning Management System is shown in figure 4.

After collecting the data from student which is stored in Moodle database, the next stage is to gather & prepare the data. In this step, first we select the data from database which is relevant. After that we clean & transformed into the format for testing.

To test the result we use the open source data mining tool WEKA. Since our project is to find the best combination of subject, we use the Apriori machine algorithm for testing the result.

In the step, build the model; we try to develop the algorithm which may be combination of various data mining algorithm. We find out the result using this model & compare the result with result obtained using already existing algorithm in WEKA. Finally we deploy the result.

4. RESULT & IMPLEMENTATION

Here we are considering the sample data extracted from Moodle database project as shown in Table2. Here we consider 45 student & 15 courses. We consider the subjects like C-programming (C), Visual Basic (VB), Active Server Pages (ASP), Computer Network (CN), Microprocessor (MP), Computer Organization (CO), Database Engineering (DBE), Advanced Database System (ADS), Operating System (OS), Distributed System (DS-I), Finite Automata System (FSA), Data Structure (DS), Software Engineering (SE), and Software Testing & Quality assurance (STQA). This is the first step in our architecture i.e. collection of data .In this table yes represent that the student is interested in that subject & no represent that student do not like that subject .

Table1 2: Prototype data assumed for architecture

Subject → Roll No of Student ↓	C	VB	ASP	CN	NE	MP	CO	DBE	ADS	OS	DS	FSA	DS-I	SE	STQA
01	yes	yes	yes	yes	yes	no	no	no	no	no	no	no	yes	no	no
02	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
03	yes	yes	yes	yes	yes	no	no	no	no	yes	yes	yes	yes	yes	yes
04	no	no	no	yes	yes	no	yes	no	no	no	no	no	no	no	no
05	yes	yes	yes	yes	yes	no	no	yes	no	yes	yes	no	yes	no	no
06	yes	yes	yes	no	no	no	no	no	no	yes	no	no	yes	no	no
07	no	no	no	yes	yes	yes	yes	no	no	no	no	no	no	yes	no
08	no	no	no	no	no	no	no	yes	yes	yes	yes	no	yes	no	no
09	no	no	no	yes	yes	yes	yes	no	no	no	no	yes	no	no	no
10	yes	no	no	no	no	no	no	no	no	no	no	no	no	no	no
11	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no
12	yes	yes	yes	yes	yes	no	no	no	no	no	no	no	no	no	no
13	no	no	no	no	no	no	no	yes	yes	yes	yes	no	yes	yes	yes
14	yes	yes	yes	yes	yes	no	no	no	no	yes	yes	no	no	no	no
15	yes	yes	yes	no	no	no	no	no	no	no	no	no	yes	no	no
16	no	no	no	yes	yes	no	no	yes	yes	yes	yes	no	yes	no	no
17	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	yes	yes
18	yes	yes	yes	no	no	no	no	no	no	no	no	no	no	no	no
19	no	no	no	yes	yes	yes	yes	yes	yes	no	no	no	no	no	no
20	yes	no	no	no	no	no	no	no	no	yes	yes	no	yes	yes	yes
21	yes	no	yes	no	no	yes	yes	no	no	yes	yes	yes	no	no	no
22	no	no	no	no	no	no	no	yes	yes	yes	yes	no	yes	no	no
23	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes	no	yes	no	no
24	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
25	no	yes	yes	no	no	yes	yes	yes	yes	yes	yes	no	no	no	no
26	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no
27	yes	yes	yes	yes	yes	no	no	no	no	no	no	no	no	no	no
28	no	no	no	yes	yes	no	no	no	no	yes	yes	no	yes	no	no
29	no	no	no	no	no	yes	yes	yes	yes	no	no	no	no	no	no
30	yes	yes	yes	yes	yes	no	no	no	no	no	no	no	no	yes	yes
31	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
32	yes	yes	yes	no	no	no	no	yes	yes	yes	yes	no	yes	no	no
33	no	no	no	yes	yes	no	no	no	no	yes	yes	no	yes	no	no
34	yes	yes	yes	no	no	no	no	no	no	no	no	no	no	no	no
35	no	no	no	no	no	no	no	no	no	yes	yes	no	no	no	no
36	no	no	no	yes	yes	no	no	no	no	no	no	no	yes	no	no
37	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no	no	no	no
38	no	no	no	no	no	no	no	no	no	yes	yes	yes	yes	yes	yes
39	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
40	no	no	no	no	no	no	no	no	no	no	no	no	no	yes	yes
41	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no

42	**no**	**no**	**no**	**yes**	**yes**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**
43	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**yes**	**yes**	**no**	**yes**	**no**	**no**
44	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**yes**
45	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**	**no**

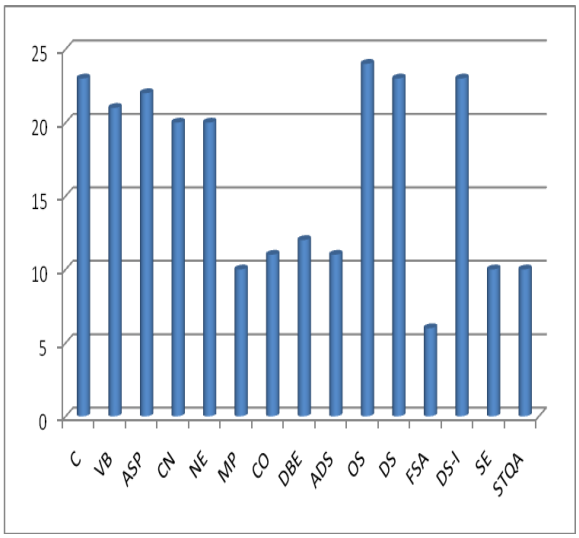

Figure 5: Subject choice before preprocessing of data

Before preprocessing, the graph for the student's subject choice is shown in figure 5. We apply the machine learning algorithm Apriori algorithm to this data. This result is shown in the 1st row in table 4.In this table, first row consist of the rule containing no only. So we need the data preprocessing. Since we are recommending the course for a student, only "yes" data is relevant to our scenario.

After preprocessing we will get the data which is given in table 3.In preprocessing, we eliminate that subject whose count is less than 20 & that student whose count is less than two. After preprocessing the data we get the data shown in table 3. After preprocessing the data, Apriori rule is applied & the result is shown in 2nd & 3rd row of table 4.

Table 3: Data after preprocessing

Subject → Roll No of Student | V	**C**	**VB**	**ASP**	**CN**	**NE**	**OS**	**DS**	**DS-I**
1	**yes**	**yes**	**yes**	**yes**	**yes**	**no**	**no**	**yes**
3	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
4	**no**	**no**	**no**	**yes**	**yes**	**no**	**no**	**no**
5	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
6	**yes**	**yes**	**yes**	**no**	**no**	**yes**	**no**	**yes**
7	**no**	**no**	**no**	**yes**	**yes**	**no**	**no**	**no**
8	**no**	**no**	**no**	**no**	**no**	**yes**	**yes**	**yes**
9	**no**	**no**	**no**	**yes**	**yes**	**no**	**no**	**no**
11	**yes**	**yes**	**yes**	**no**	**no**	**yes**	**yes**	**yes**
12	**yes**	**yes**	**yes**	**yes**	**yes**	**no**	**no**	**no**
13	**no**	**no**	**no**	**no**	**no**	**yes**	**yes**	**yes**
14	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**no**
15	**yes**	**yes**	**yes**	**no**	**no**	**no**	**no**	**yes**
16	**no**	**no**	**no**	**yes**	**yes**	**yes**	**yes**	**yes**
17	**yes**	**yes**	**yes**	**no**	**no**	**yes**	**yes**	**yes**
18	**yes**	**yes**	**yes**	**no**	**no**	**no**	**no**	**no**
19	**no**	**no**	**no**	**yes**	**yes**	**no**	**no**	**no**
20	**yes**	**no**	**no**	**no**	**no**	**yes**	**yes**	**yes**
21	**yes**	**no**	**yes**	**no**	**no**	**yes**	**yes**	**no**
22	**no**	**no**	**no**	**no**	**no**	**yes**	**yes**	**yes**
23	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
24	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
25	**no**	**yes**	**yes**	**no**	**no**	**yes**	**yes**	**no**
26	**yes**	**yes**	**yes**	**no**	**no**	**yes**	**yes**	**yes**
27	**yes**	**yes**	**yes**	**yes**	**yes**	**no**	**no**	**no**
28	**no**	**no**	**no**	**yes**	**yes**	**yes**	**yes**	**yes**
30	**yes**	**yes**	**yes**	**yes**	**yes**	**no**	**no**	**no**
32	**yes**	**yes**	**yes**	**no**	**no**	**yes**	**yes**	**yes**
33	**no**	**no**	**no**	**yes**	**yes**	**yes**	**yes**	**yes**
34	**yes**	**yes**	**yes**	**no**	**no**	**no**	**no**	**no**
35	**no**	**no**	**no**	**no**	**no**	**yes**	**yes**	**no**
36	**no**	**no**	**no**	**yes**	**yes**	**no**	**no**	**yes**
37	**yes**	**yes**	**yes**	**yes**	**yes**	**no**	**no**	**no**
38	**no**	**no**	**no**	**no**	**no**	**yes**	**yes**	**yes**
39	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**	**yes**
41	**yes**	**yes**	**yes**	**no**	**no**	**yes**	**yes**	**yes**
42	**no**	**no**	**no**	**yes**	**yes**	**no**	**no**	**no**
43	**no**	**no**	**no**	**no**	**no**	**yes**	**yes**	**yes**

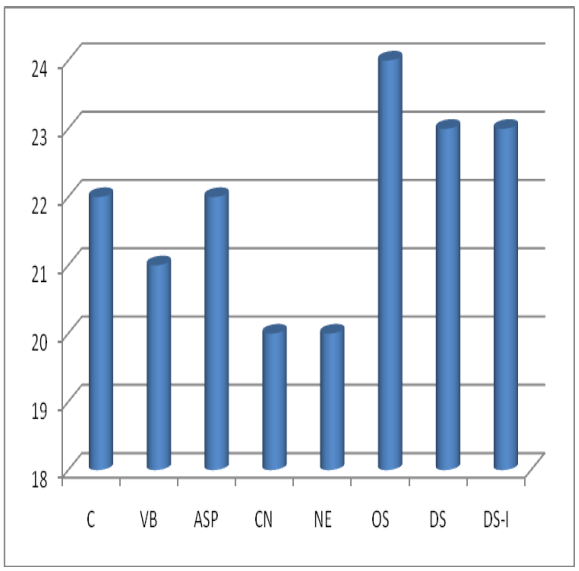

Figure 6: Subject choice after preprocessing of data

Graph for student's subject choice after preprocessing of data is shown in figure 6. After applying the Apriori Algorithm which is already existing algorithm in the WEKA tool, we get the following result regarding the best combination of subjects as shown in 2nd & 3rd row in table 4. In 2nd row, lower support bound is 0.5 & upper support bound is 1.0. If we increase the lower support bound to 0.6 then we get the refined rule shown in 3rd row of table 4. The meaning of the rule is: if student is interested in Distributed system then he/she is interested in Operating System & vice versa. Due to above rule, we can recommend to new student who has recently enrolled for DS course, the Operating System as a course to be opted. Association rule obtained here also match with the in general real world interdependencies among the course.

Table 4: Result of applying machine learning algorithm: Apriori Rule

Sr. No.	Courses considered	Parameter considered	Result using machine learning algorithm e.g. Apriori Association Algorithm
Before preprocessing of data			
1.	1.C Programming 2.Visual Basic 3.Active Server Pages 4.Computer Network 5.Network Engineering 6.Microprocessor 7.Computer Organization 8.Database Engineering 9.Advanced Database System 10.Oerating System 11.Distributed System 12.Finite State Automata 13.Data Structure-I 14.Software Engineering 15.Software Testing & Quality Assurance	1.Minimum support: 0.7 2.Minimum metric <confidence>: 0.9 3.Number of cycles performed: 6	1. Computer_Organization=no → Microprocessor=no 2.Database_Engineering=no → Advanced_Database_System=no 3.Computer_Organization=no Finite_State_Automata=no → Microprocessor=no 4.Microprocessor=no → Computer_Organization=no 5.Softare_Tesing_And_Quality_Assurance=no → Software_Engineering=no 6.Software_Engineering=no → Softare_Tesing_And_Quality_Assurance=no 7.Advanced_Database_System=no → Database_Engineering=no 8.Microprocessor=no Finite_State_Automata=no → Computer_Organization=no 9.Finite_State_Automata=no Softare_Tesing_And_Quality_Assurance=no → Software_Engineering=no 10. Finite_State_Automata=no Software_Engineering=no → Softare_Tesing_And_Quality_Assurance=no
After preprocessing of data			
2.	1.C Programming 2.Visual Basic 3.Active Server Pages 4.Computer Networks 5.Network Engineering 6.Operating System 7.Distributed Systems 8.Data Structures	1.Minimum support: 0.5 2.Minimum metric <confidence>: 0.9 3.Number of cycles performed: 10	1. Distributed_Systems=yes → Operating_System=yes 2. Visual_Basic=yes → Active_Server_Pages=yes 3. Network_Engineering=yes → Computer_Networks=yes 4. Computer_Networks=yes → Network_Engineering=yes 5. C_Programming=yes Visual_Basic=yes → Active_Server_Pages=yes 6. Distributed_Systems=yes Data_Structure=yes → Operating_System=yes 7. Operating_System=yes → Distributed_Systems=yes 8. Active_Server_Pages=yes → C_Programming=yes 9. C_Programming=yes → Active_Server_Pages=yes 10.Active_Server_Pages=yes → Visual_Basic=yes
3.	1.C Programming 2.Visual Basic 3.Active Server Pages 4.Computer Networks 5.Network Engineering 6.Operating System 7.Distributed Systems 8.Data Structures	1.Minimum support: 0.6 2.Minimum metric <confidence>: 0.9 3.Number of cycles performed: 8	1. Distributed_Systems=yes → Operating_System=yes 2.Operating_System=yes → Distributed_Systems=yes

5. CONCLUSION AND FUTURE WORK

In this paper we are trying to find out the best combination of courses which will help us to study the behavior of student regarding the courses i.e. in which subjects students are interested. After collecting the sample data from Moodle database, we try to find out the combination of machine learning algorithms which are better at predicting course selection by the student. In our preliminary stage, we find that association rule e.g. Apriori association rule which is existing machine learning algorithm in open source data mining tool WEKA, is good at obtaining the prediction. Future work includes obtaining the results i.e. the best combination of subjects, for the combination of various data mining algorithms.

6. REFERENCES

[1] Castro, F., Vellido, A., Nebot, A., & Mugica, F. (in press). Applying data mining techniques to e-learning problems: A survey and state of the art. In L. C. Jain, R. Tedman, & D. Tedman (Eds.), Evolution of

Teaching and learning paradigms in intelligent environment. Studies in Computational Intelligence (Vol. 62). Springer-Verlag.

[2] C. Romero, S. Ventura and E. Garcia. Data Mining in Course Management Systems: MOODLE Case Study and Tutorial. Computers and Education, 2007. Num. 51. pp. 368-384.

[3] C. Carmona, G. Castillo and E. Millán: Discovering Student Preferences in E-learning, EC-TEL07, pp.33-42 (2007)

[4] Hamalainen, W., Suhonen, J., Sutinen, E., & Toivonen, H. (2004). Data mining in personalizing distance education courses. In World conference on open learning and distance education, Hong Kong (pp. 1–11).

[5] Zaı¨ane, O. (2002). Building a recommender agent for e-learning systems. In Proceedings of the international conference in education, Auckland, New Zealand (pp. 55–59).

[6] Resende, S.D., Pires, V.M.T.: Using Data Warehouse and Data Mining Resources for Ongoing Assessment of Distance Learning. In: IEEE International Conference on Advanced Learning Technologies, ICALT 2002. (2002).

[7] "Data Mining Introductory and Advanced Topics" by Margaret H. Dunham

[8] Weka (2007). http://www.cs.waikato.ac.nz/ml/weka/.

[9] International Educational Data Mining, available at http://www.educationaldatamining.org/

[10] Sunita B Aher and Lobo L.M.R.J.. Data Mining in Educational System using WEKA. *IJCA Proceedings on International Conference on Emerging Technology Trends (ICETT)* (3):20-25, 2011. Published by Foundation of Computer Science, New York, USA (ISBN: 978-93-80864-71-13)

[11] Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás: Data Mining Algorithms to Classify Students

7. AUTHORS PROFILE

Sunita B Aher received the B.E degre in Computer Science & Engineering in 2000 from Amravati University, Amravati, India and pursuing the M. E. degree in Computer Science in Walchand Institute of Technology, Solapur, India.

She is doing the dissertation work under the guidance of Mr. Lobo l.M.R.J, Associate Professor & Head, Department of IT, Walchand Institute of Technology, Solapur, Maharashtra, India.

Mr. **Lobo l.M.R.J,** Associate Professor & Head, Department of IT, Walchand Institute of Technology, Solapur, Maharashtra, India.

He has registered for Ph.D in Computer Science and Engineering at SGGS, Nanded of Sant Ramanand Teerth Marathawada University, and Nanded, India. Under the guidance of Dr. R.S. Bichkar. He is presently working as an Associate Professor & Head, Department of IT Walchand Institute of Technology, Solapur, Maharashtra, India. His research interests include Evolutionary Computation, Genetic Algorithms and Data Mining.