Association Rule Mining of Classified and Clustered Data of 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
Classification is the supervised learning algorithm that maps the data into predefined groups & classes. Clustering is unsupervised algorithm which finds groups of objects such that the objects in one group will be similar to one another and different from the objects in another group while Association Rule algorithms are used to show the relationship between the data items. In this paper we propose the combination of three data mining algorithms: ADTree classification algorithm, Simple K-means Clustering Algorithm & Apriori Association Rule algorithm to recommend the course selected by the student. Here we consider the real sample data of Moodle courses of our college & check the result using the open source data mining tool Weka. First we classify the data using ADTree classification algorithm & then we apply the Simple k-means algorithm to the resultant data to obtain clusters. We apply the Apriori Association Rule algorithm on clusters obtained to find the best combination of courses which gives the better result as compare to result we obtained using only the Apriori Association Rule.

KEYWORDS
Weka, ADTree Classification Algorithm, Simple K-means Algorithm, Apriori Association Rule Algorithm, Learning management System

1. INTRODUCTION
Data mining can also be used to extract the knowledge from E-learning system such as Moodle. The course recommendation system in e-learning is a system that suggests the best combination of courses in which the students are interested [7]. For collecting the data we consider the student of three years of engineering course i.e. Second year, Third year, & Final year of Computer Science & Engineering and Information Technology. Here we are using Moodle as Learning Management System to collect the data regarding the course selection by student. We have created the student login & gave the access to the student. The data flow diagram (DFD) for Recommendation of Courses in E-Learning System is shown in figure 1. In this Course Recommendation System, we have considered the 13 course category. Under each category there will courses. So there are about 82 courses all together.

In this DFD , 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 which they would like to. This enrollment information is stored in database [10].

After collecting the data from student which is stored in Moodle database, the next stage is to select the relevant data from Moodle database. After selecting the data next stage is to preprocess it which is explained in paper [9]. 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.

Now we consider the combined approach i.e. classification, clustering & association rule algorithm. First we apply the classification algorithm ADTree to the data selected from Moodle database. After classifying the data, we apply the Simple K-means clustering algorithm on this classified data. We apply the Apriori association rule algorithm to this clustered data to find the association rule. We compare the result of combined approach to the result obtained using only the Apriori Association Rule algorithm. If we use the combined approach then there is no need to preprocess the data.

2. LITERATURE REVIEW
In paper [1], they focused on the cases where the items of a large domain correlate with each other in a way that small worlds are formed, that is, the domain is clustered into groups with a large number of intra-group and a small number of inter-group correlations. This property appears in several real-world cases, e.g., in bioinformatics, e-commerce applications, and bibliographic analysis, and can help to significantly prune the search space so as to perform efficient association-rule mining. They developed an algorithm that partitions the domain of items according to their correlations and they describe a mining algorithm that carefully combines partitions to improve the efficiency. Their experiments show the superiority of the proposed method against existing algorithms, and that it overcomes the problems (e.g., increase in CPU cost and possible I/O thrashing) caused by existing algorithms due to the combination of a large domain and a large number of records.

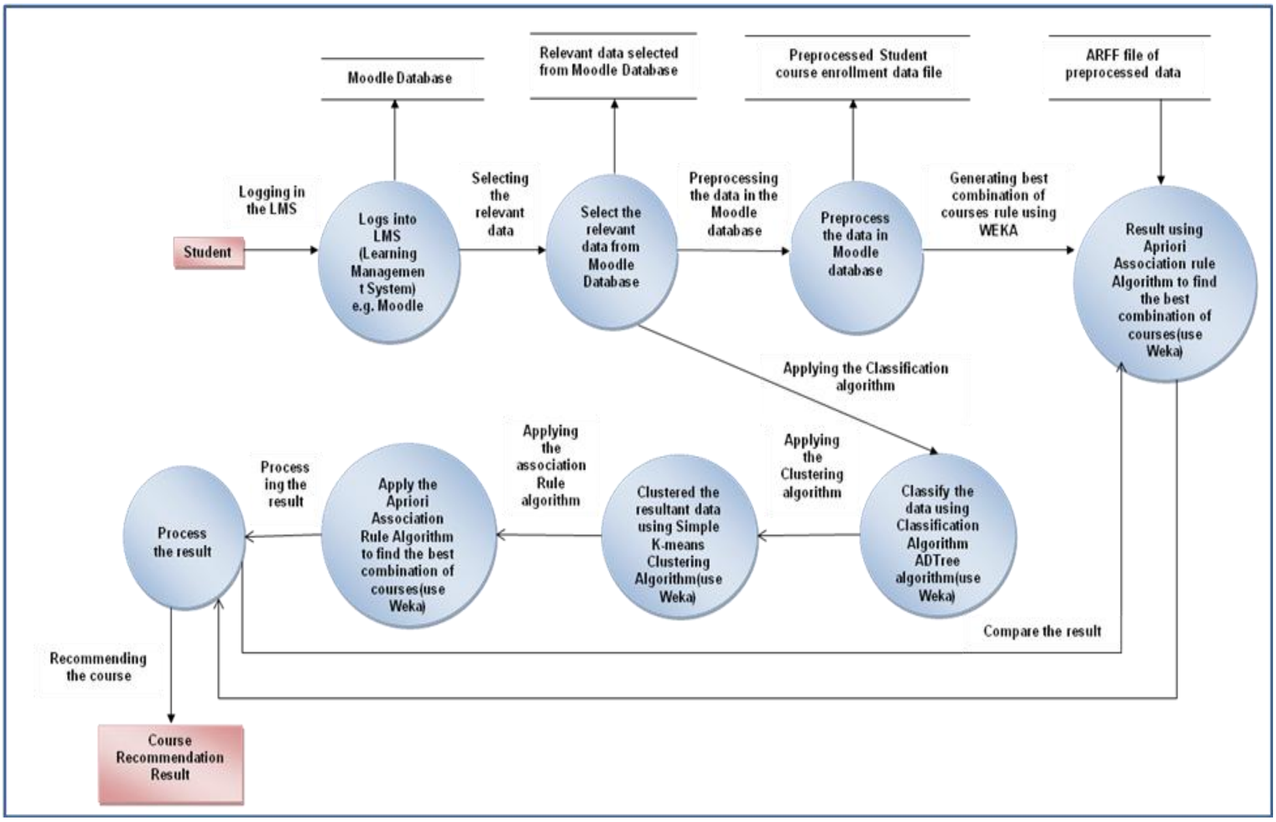

Figure 1: Data flow diagram for recommendation of courses in E-learning system

The research [2] compares the performance of three popular association rule mining algorithms, namely Apriori, predictive Apriori and Tertius based on data characteristics. The accuracy measure is used as the performance measure for ranking the algorithms. A meta-learning technique is implemented for a unique selection from a set of association rule mining algorithms. On the basis of experimental results of 15 UCI data sets, this research discovers statistical information based rules to choose a more effective algorithm.

In paper [3], they conducted experiment in the WEKA environment by using four algorithms namely ID3, J48, Simple CART and Alternating Decision Tree on the spam email dataset and later the four algorithms were compared in terms of classification accuracy. According to their simulation results the J48 classifier outperforms the ID3, CART and ADTree in terms of classification accuracy.

The research [4] analyzed alternative measures which could replace confidence in order to evaluate the suitability of a given association rule with respect to the classification problem when building a classification model.

In paper [5], they proposed a approach called Associative classification which is a classification of a new tuple using association rules. It is a combination of association rule mining and classification. The accuracy can be achieved by producing all types of negative class association rules.

In paper [6], they explain a novel aspect mining method which combines clustering and association rule technology Clustering analysis based on the execution traces is provided to find out candidate aspects; while association rule mining based on the execution traces with ordered call is used to find out the crosscuts. Both the aspect code (advice body) and the crosscuts (pointcuts) are gotten after the above two processes, which constitute the aspect mining process.

3. DATA MINING ALGORITHMS

Here we consider brief idea about each data mining algorithm: ADTree classification algorithm, Simple K-means clustering algorithm & Apriori association rule algorithm.

3.1 ADTree Classification Algorithm

An alternating decision tree (ADTree) is a machine learning method for classification which generalizes decision trees.

An alternating decision tree consists of two nodes. Decision nodes specify a predicate condition & prediction nodes contain a single number. ADTree always have prediction nodes as both root and leaves. The inputs to the alternating
decision tree algorithm is a set of inputs (x1, y1),…..,(xm,ym) where x_i is a vector of attributes and y_i is either -1 or 1 and a set of weights w_i corresponding to each instance. Inputs are also called instances.

An instance is classified by an ADTree by following all paths for which all decision nodes are true and summing any prediction nodes that are traversed.

3.2 Simple K-means Algorithm

Simple K-means algorithm is unsupervised algorithm in which items are moved among the set of cluster until required set is reached [7].

The steps for Simple K-means algorithms are:

Step 1: Define k centroids, one for each cluster.

Step 2: Take each point belonging to a given data set and associate it to the nearest centroid till no

point is pending.
Step 3: Recalculate k new centroids
Step 4: Repeat step 2 & 3 until there are no changes in centroid.

3.3 Apriori Association Rule Algorithm

Apriori Association rule is used to mine the frequent patterns in database. 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. 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].
Apriori Association Rule algorithm is shown in figure 2.

Input : Database of Transactions D={t_1,t_2, …, t_n}
Set if Items I= {I_1 ,I_2 ,….I_k}
Frequent (Large) Itemset L
Support,
Confidence.
Output : Association Rule satisfying Support & Confidence
Method :

1. C_1 = Itemsets of size one in I;
2. Determine all large itemsets of size 1, L_1;
3. i = 1;
4. Repeat
5. i = i + 1;
6. C_i = Apriori-Gen(L_{i-1});
7. Apriori-Gen(L_{i-1})
    1. Generate candidates of size i+1 from large itemsets of size i.
    2. Join large itemsets of size i if they agree on i-1.
    3. Prune candidates who have subsets that are not large.
8. Count C_i to determine L_i;
9. until no more large itemsets found;

Figure 2: Apriori Association Rule algorithm [7]

Figure 4 shows the generation of itemsets & frequent itemsets where the minimum support count is 3

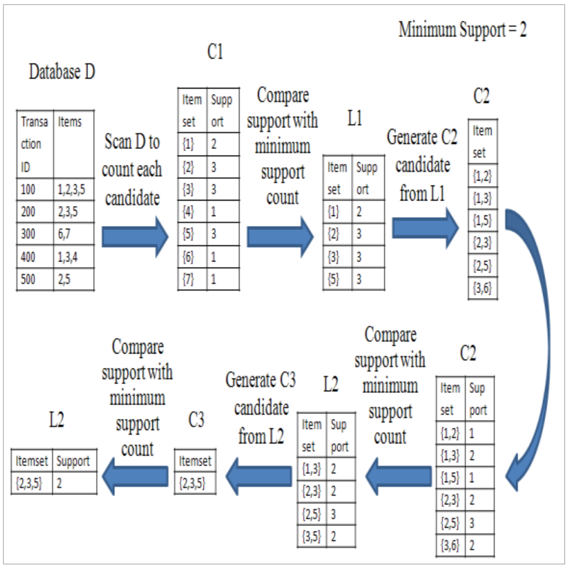

Figure 3: Generation of itemsets & frequent itemsets [8]

To generate the association rule from frequent itemset we use the following rule:

- For each frequent itemset L, find all nonempty subset of L
- For each nonempty subset of L, write the association rule S → (L-S) if support count of L/support count of S >= Minimum Confidence

The best rule from the itemset L= {2, 3, 5} are calculated as follows:
Consider the minimum support is 2 & minimum confidence is 70%. All nonempty subset of {2,3,5} are: {2,3},{2,5},{3,5},{2},{3},{5}.
Rule 1: {2, 3} → {5}
Confidence = Support Count of ({2, 3, 5})/ Support Count of ({2, 3}) = 2/2 = 100%
Rule 2: {2, 5} → {3}
Confidence = Support Count of ({2, 3, 5})/ Support Count of ({2, 5}) = 2/3 = 67%
Rule 3: {3, 5} → {2}
Confidence = Support Count of ({2, 3, 5})/ Support Count of ({3, 5}) = 2/2 = 100%
Rule 4: {2} → {3, 5}
Confidence = Support Count of ({2, 3, 5})/ Support Count of ({2}) = 2/3 = 67%
Rule 5: {3} → {2, 5}
Confidence = Support Count of ({2, 3, 5})/ Support Count of ({3}) = 2/3 = 67%
Rule 6: {5} → {2, 3}
Confidence = Support Count of ({2, 3, 5})/ Support Count of ({5}) = 2/3 = 67%
Hence the accepted rules are Rule 1 & Rule 3 as the confidence of these rules is greater than 70%.

4. RESULT & IMPLEMENTATION

Here we are considering the sample data extracted from Moodle database as shown in Table 1. Here we consider 45 student & 15 courses. We consider the courses like C-programming (C), Visual Basic (VB), Active Server Pages (ASP), Computer Network (CN), Network Engineering (NE), Microprocessor (MP), Computer Organization (CO), Database Engineering (DBE), Advanced Database System (ADS), Operating System (OS), Distributed System (DS), Finite Automata System (FSA), Data Structure (DS-I), Software Engineering (SE), and Software Testing & Quality assurance (STQA). In this table, "yes" represent that the student is interested in that course & "no" represent that student do not like that course .First we preprocess the data obtained from Moodle course, to remove the noise or to add the missing value, while constructing the sample table which is given in table 1. In our preprocessing step [9], we delete those rows & columns from sample table shown in table 1, having very less student count & less course count. After preprocessing of data, we got 8 courses & 38 students. These 8 courses are C-programming (C), Visual Basic (VB), Active Server Pages (ASP), Computer Network (CN), Network Engineering (NE), Operating System (OS), Distributed System (DS), Data Structure (DS-I). The graph for sample data before preprocessing & after preprocessing is shown in figure 4 & 5 respectively. In these graph the X-axis represents the courses & Y-axis represents the course count.
The sample table 1 after preprocessing of data is shown in table 2. The result of applying Apriori association rule before & after preprocessing of data is shown in first & second row of table 5.

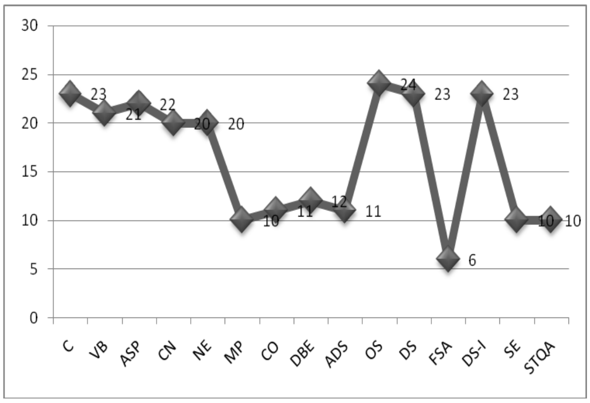

Figure 4: Graph for sample table 1 from Moodle courses before preprocessing

Before preprocessing of data, we got the association rule containing “no” only. As we are recommending the course, we preprocess the data. The result after preprocessing of data is shown in second row of table 5. Now the association rule contains only “yes”. The meaning of the association rule “DS=yes -> OS=yes” is that we can recommend to new student who has recently enrolled for DS course, the operating system as a course to be opted.

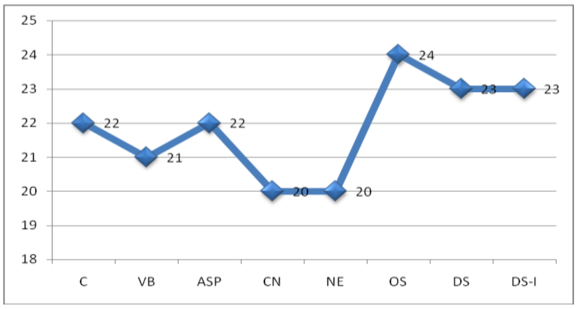

Figure 5: Graph for sample table 1 after preprocessing

If we consider combination of classification, clustering & association rule then there is no need to preprocess the data [9]. First we apply the ADTree classification algorithm to data selected from Moodle database. After classification of data, we got the table 3. After classifying the data, we apply the Simple K-means clustering algorithm on classified data & we got the table which is shown in table 4. While applying the clustering algorithm to classified data, we consider the three clusters out of which only the cluster 0 gives the correct result. Cluster 2 in clustered data need not be considered since it will definitely give the correct result. Last step is to apply the association rule to this clustered data. The result of application of Apriori association rule to this classified & clustered data is shown in fourth row of table 5.

Table 1: Sample table from Moodle Database

Courses ➔ Roll_No | v	C	VB	ASP	CN	NE	MP	CO	DBE	ADS	OS	DS	FSA	DS-I	SE	STQA
1	yes	yes	yes	yes	yes	no	no	no	no	no	no	no	yes	no	no
2	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
3	yes	yes	yes	yes	yes	no	no	no	no	yes	yes	yes	yes	yes	yes
4	no	no	no	yes	yes	no	yes	no	no	no	no	no	no	no	no
5	yes	yes	yes	yes	yes	no	no	yes	no	yes	yes	no	yes	no	no
6	yes	yes	yes	no	no	no	no	no	no	yes	no	no	yes	no	no
7	no	no	no	yes	yes	yes	yes	no	no	no	no	no	no	yes	no
8	no	no	no	no	no	no	no	yes	yes	yes	yes	no	yes	no	no
9	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

Table 2: Sample table 1 after preprocessing of data

Sub ject ➔ Roll No of Student | v	C	VB	AS P	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

Table 3: After application of classification algorithm to table1

Courses ➔ Roll_No | v	C	VB	ASP	CN	NE	MP	CO	DBE	ADS	OS	DS	FSA	DS-I	SE	STQ A
3	yes	yes	yes	yes	yes	no	no	no	no	yes	yes	yes	yes	yes	yes
5	yes	yes	yes	yes	yes	no	no	yes	no	yes	yes	no	yes	no	no
11	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no
17	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	yes	yes
20	yes	no	no	no	no	no	no	no	no	yes	yes	no	yes	yes	yes
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
26	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no
32	yes	yes	yes	no	no	no	no	yes	yes	yes	yes	no	yes	no	no
39	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
41	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no

Table 4: After applying the clustering technique on table 3

Courses ➔ Roll_No | v	C	VB	ASP	CN	NE	MP	CO	DBE	ADS	OS	DS	FSA	DS-I	SE	ST QA
5	yes	yes	yes	yes	yes	no	no	yes	no	yes	yes	no	yes	no	no
11	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no
23	yes	yes	yes	yes	yes	yes	yes	no	no	yes	yes	no	yes	no	no
26	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no
32	yes	yes	yes	no	no	no	no	yes	yes	yes	yes	no	yes	no	no
41	yes	yes	yes	no	no	no	no	no	no	yes	yes	no	yes	no	no

Table 5: Result after application of machine learning algorithms

Course considered		**Results**
Result of Apriori Association Rule before preprocessing & application of combination of Clustering & Association Rule		
C, VB, ASP, CN, NE, MP, CO, DBE, ADS, OS, DS, FSA, DS-I, SE, STQA	Minimum support: 0.7 Minimum metric <confidence>: 0.9	Best rules found: 1. CO=no → MP=no conf:(1) 2.DBE=no → ADS=no conf:(1) 3.CO=no FSA=no → MP=no conf:(1) 4.MP=no → CO=no conf:(0.97) 5.STQA=no → SE=no conf:(0.97) 6.SE=no → STQA=no conf:(0.97) 7.ADS=no → DBE=no conf:(0.97) 8.MP=no FSA=no →CO=no conf:(0.97) 9.FSA=no STQA=no → SE=no conf:(0.97) 10. FSA=no SE=no → STQA=no conf:(0.97)
Result of Apriori Association Rule after preprocessing & before application of combination of Clustering & Association Rule		
C, VB, ASP, CN, NE, OS, DS, DS-I	Minimum support: 0.5 Minimum metric <confidence>: 0.9	Best rules found: 1. DS=yes →OS=yes conf:(1) 2. VB=yes →ASP=yes conf:(1) 3. NE=yes → CN=yes conf:(1) 4. CN=yes → NE=yes conf:(1) 5. C=yes VB=yes → ASP=yes conf:(1) 6. DS=yes DS-I=yes → OS=yes conf:(1) 7. OS=yes →DS=yes conf:(0.96) 8. ASP=yes →C=yes conf:(0.96) 9. C=yes →ASP=yes conf:(0.96) 10.ASP=yes →VB=yes conf:(0.96)
	Minimum support: 0.6 Minimum metric <confidence>: 0.9	Best rules found: 1. DS=yes →OS=yes conf:(1) 2.OS=yes →DS=yes conf:(0.96)
After Application of Classification algorithm-ADTree & Clustering algorithm-Simple K-means Algorithm		
C, VB, ASP, CN, NE, MP, CO, DBE, ADS, OS, DS, FSA, DS-I, SE, STQA	Number of Cluster:3 Seed: 10	Cluster 0 Mean/Mode: 5.5 yes yes yes no no no no no no yes yes no yes no no Std Devs: 3.5071 N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A Cluster 1 Mean/Mode: 2.3333 yes yes yes no no no no no no yes yes no yes yes yes Std Devs: 2.0817 N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A Cluster 2 Mean/Mode: 7.5 yes yes yes yes yes yes yes yes yes yes yes yes yes yes yes Std Devs: 2.1213 N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A N/A Clustered Instances 0 6 (55%) 1 3 (27%) 2 2 (18%)
After Application of Classification algorithm-ADTree , Clustering algorithm-Simple K-means Algorithm & Association Rule-Apriori Association Rule		

C, VB, ASP, CN, NE, MP, CO, DBE, ADS, OS, DS, FSA, DS-I, SE, STQA	Minimum support: 0.95 Minimum metric <confidence>: 0.9	Cluster 0 (Correct Result) Best rules found: 1. VB=yes → C=yes conf:(1) 2. C=yes → VB=yes conf:(1) 3. ASP=yes → C=yes conf:(1) 4. C=yes → ASP=yes conf:(1) 5. OS=yes → C=yes conf:(1) 6. C=yes → OS=yes conf:(1) 7. DS=yes → C=yes conf:(1) 8. C=yes → DS=yes conf:(1)
	Minimum support: 0.95 Minimum metric <confidence>: 0.9	Cluster 1 (Incorrect Result) Best rules found: 1. MP=no3 → C=yes conf:(1) 2. C=yes3 → MP=no conf:(1) 3. CO=no → C=yes conf:(1) 4. C=yes → CO=no conf:(1) 5. DBE=no → C=yes conf:(1) 6. C=yes → DBE=no conf:(1) 7. ADS=no → C=yes conf:(1) 8. C=yes → ADS=no conf:(1)

The graph for courses after application of ADTree classification algorithm on Table 1 is shown in figure 6.

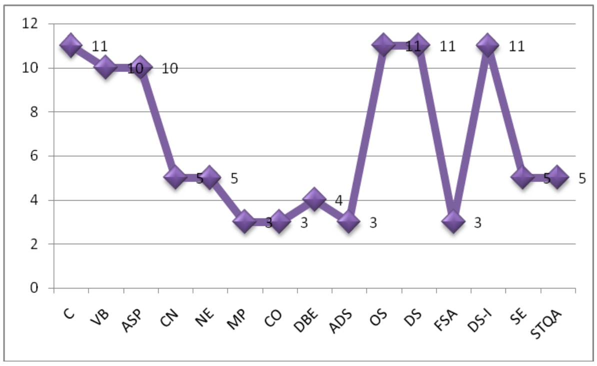

Figure 6:Graph after application of ADTree classification algorithm on Table 1

After application of clustering algorithm on classified data we got two cluster out of which the cluster 0 is the correct cluster & cluster 1 is incorrect. The graph for cluster 0 & cluster 1 after application of clustering algorithm on classified data ia shown in figure 7 & 8 respectively.

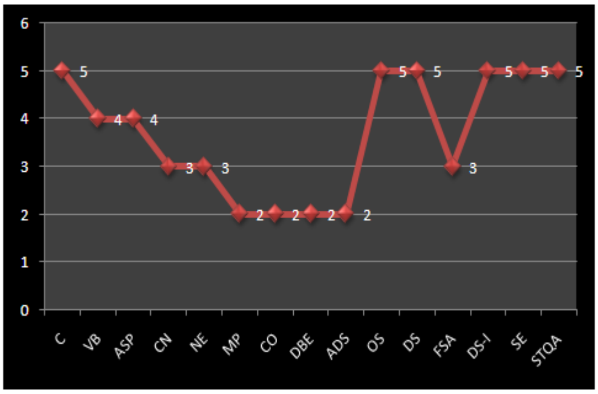

Figure 7:Graph for cluster 1 after application of Simple K-means clustering algorithm on classified data table 3
(incorrect result using cluster 1)

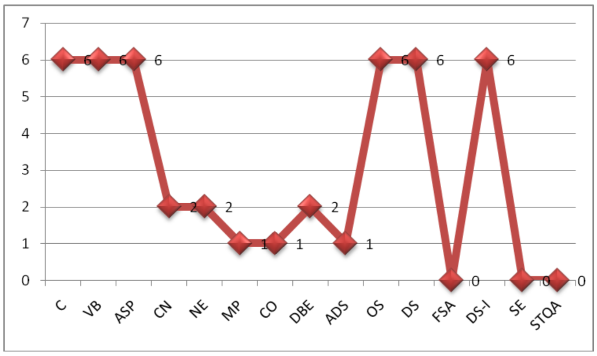

Figure 8:Graph for cluster 0 after application of Simple K-means clustering algorithm on classified data table 2
(correct result using cluster 0)

The overall graph is shown in figure 9.The line ST, ADT, ADTSKM0 Correct & ADTSKM1 in graph represents courses considered in sample table, after application of ADTree classification algorithm, after application of Simple K-means clustering algorithm on classified data (cluster 0-correct cluster) & after application of Simple K-means clustering algorithm on classified data (cluster 1-incorrect cluster) respectively.

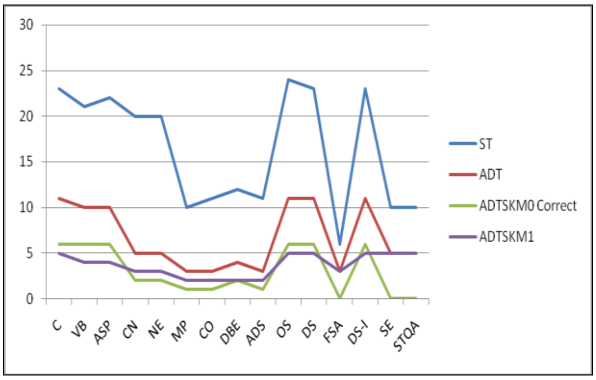

Figure 9: Graph for courses after application of various data mining algorithms

5. COMPARATIVE STUDY OF COMBINED APPROACH & APRIORI ASSOCIATION RULE

With the result using only Apriori association rule algorithm, as we increase the support then we get the refined rule but the number of rules, we get, are less which is shown in second row of table 5. For support 0.6 we get two association rules only. If we use the combined approach i.e.combination of classification, clustering & association rule algorithm then we get eight association rules for support 0.95 which is shown in fourth row of table 5.

6. CONCLUSION AND FUTURE WORK

Classification, Clustering & Association rule are three algorithms in data mining which we are using to find out the best combination of courses in E-Learning System. Here in this paper, we consider & propose the combination of these three algorithms to recommend the course to the student. We preprocess the data obtained from Moodle courses for removing the noise or adding the missing value while constructing the final sample table. We compare the result of this combined approach with the result obtained using only the Apriori association rule algorithm. According to our simulation the result of this combined approach work better than only Apriori association rule algorithm because with Apriori association rule algorithm, if we increased the support then we get the refined rule but the number of rules are less & also we have to preprocess the data which is explained [9]. Future work includes the atomization of this combined approach& to test result on huge amount of data obtained from the Moodle courses of a college.

7. REFERENCES

[1] Alexandros Nanopoulos, Apostolos N. Papadopoulos, Yannis Manolopoulos:" Mining association rules in very large clustered domains" Information Systems 32 (2007) 649–669

[2] Mazid, M.M. Shawkat Ali, A.B.M. Tickle, K.S. : "Finding a unique Association Rule Mining algorithm based on data characteristics"

[3] Aman Kumar Sharma, Suruchi Sahni:" A Comparative Study of Classification Algorithms for Spam Email Data Analysis" in International Journal of computer science & Engineering.

[4] Daniel Sánchez, Jose-María Serrano, Amparo Vila, Fernando Berzal, Juan-Carlos Cubero, Nicolás Marín:"Association rule evaluation for classification purposes" TAMIDA2005, pp.135-144 ISBN: 84-9732-449-8 © 2005

[5]B.Ramasubbareddy, A.Govardhan & A.Ramamohanreddy:"Classification Based on Positive and Negative Association Rules" International Journal of Data Engineering, (IJDE), Volume (2): Issue (2) : 2011 84

[6] *Lili He, Hongtao Bai:"*Aspect Mining Using Clustering and Association Rule Method" IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.2A, February 2006

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

[8] 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)

[9] Sunita B Aher and Lobo L.M.R.J.: "Preprocessing Technique for Association Rule Based Course Recommendation System in E-learning" selected in ICECT-2012, proceeding will published by IEEE.

[10] Sunita B Aher and Lobo L.M.R.J. Article: A Framework for Recommendation of courses in E-learning System. *International Journal of Computer Applications* 35(4):21-28, December 2011. Published by Foundation of Computer Science, New York, USA **ISSN 0975 – 8887**