

Mining Association Rule in Classified Data for Course Recommender System in E-learning

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 ADTree (Alternating Decision Tree) is a supervised classification technique that combines decision trees with the predictive accuracy into a set of classification rules & association rule algorithms are used to show the relationship between data items. Here in this paper we combine these two algorithms & apply it to sample data obtained from Moodle courses of our college for the Course Recommender System which predicts the course selected by the students. First we consider the result using only the association rule then we consider this combined approach. Here we present the advantage of applying the combined approach to Course Recommender System as compare to the result of applying only the association rule algorithm. We found that combined approach works better than only the association rule mining. This combined approach also increases the strength of the rules.

KEY WORDS

ADTree Classification algorithm, Apriori Association Rule algorithm, Weka

1. INTRODUCTION

Data mining also known as Knowledge Discovery in Database (KDD) is the extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data. Alternative names to data mining are knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc. [8].

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 Course Recommendation System is shown in figure 2. In this Course Recommendation System, we have considered the 13 course category. Under each category there will courses. So there are about 82 courses.

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 in which they like. This enrollment information is stored in database [7].

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 & association rule algorithm. First we apply the classification algorithm ADTree to the data selected from Moodle database. After classifying the data, we apply the Apriori Association Rule algorithm to find the result. 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 studied a new technique called post-bagging, which consists in resampling parts of a classification model rather than the data. They did this with a particular kind of model: large sets of classification association rules, and in combination with ordinary best rule and weighted voting approaches. They empirically evaluated the effects of the technique in terms of classification accuracy.

In paper [2], a novel combination strategy for multi-class classification (CSMC) based on multiple rules is proposed. In CSMC, rules are regarded as classification experts, after the calculation of the basic probability assignments (bpa) and evidence weights; Yang's rule of combination is employed to combine the distinct evidence bodies to realize an aggregate classification. The comparison with popular methods like CBA, C4.5, RIPPER and MCAR indicates that CSMC is a competitive method for classification based on association rule.

In paper [3], they have explored and applied the combinatorial mathematics in class association rule mining. Their algorithm is based on producing combinations of itemsets including classes using combinatorial mathematics and subsequently finds an associative classifier. In their approach, experimental results have given accuracy, near to other popular classification methods.

In paper [4], they proposed Associative classification which is a classification of a new tuple using association rules. It is a combination of association rule mining and classification. They searched for strong associations between frequent patterns and class labels. The main aim of this paper is to improve accuracy of a classifier. The accuracy can be achieved by producing all types of negative class association rules.

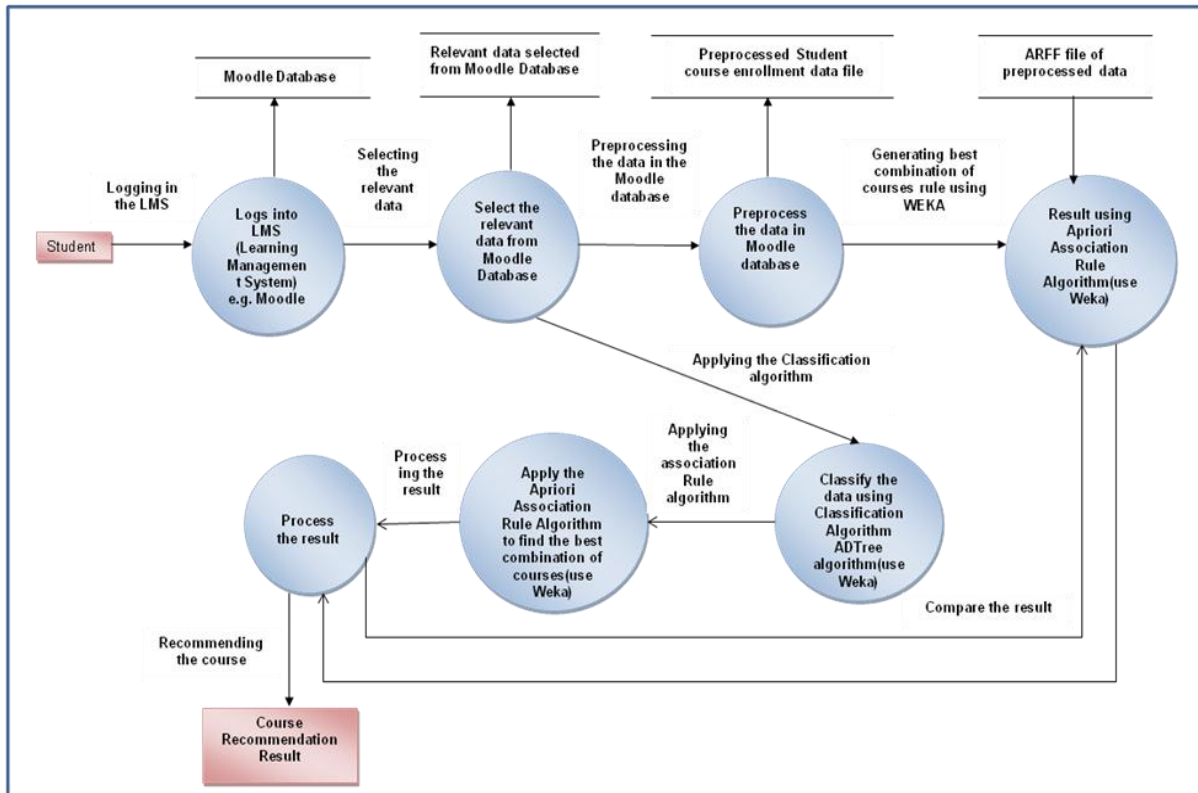


Figure 1: Data flow diagram for Course Recommender System

In paper [5], they proposed a novel aspect mining method which combines clustering and association rule technology is provided in this article. 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.

The paper [10] analyzed alternative measures which could replace confidence in order to evaluate the suitability of a given association rule with respect to the classification problem they tried to solve when building a classification model.

In the paper [11], classification and association rule mining algorithms are discussed and demonstrated. Particularly, the problem of association rule mining, and the investigation and comparison of popular association rules algorithms. The classic problem of classification in data mining was also discussed. The paper also considered the use of association rule mining in classification approach in which a recently proposed algorithm was demonstrated. Finally, a comprehensive experimental study against 13 UCI data sets is presented to evaluate and compare traditional and association rule based classification techniques with regards to classification accuracy, number of derived rules, rules features and processing time.

3. 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.

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.

The inputs to the alternating decision tree algorithm are:

- Inputs are also called instances. A set of inputs $(x_1, y_1), \dots, (x_m, y_m)$ where x_i is a vector of attributes and y_i is either -1 or 1. Inputs are also called instances and
- A set of weights w_i corresponding to each instance.

The fundamental element of the ADTree algorithm is the rule as it comes under the tree category. A single rule consists of [12]:

- Condition: is a predicate of the form "attribute <comparison> value.
- Precondition: is simply a logical conjunction of conditions and
- Two scores.

A precondition Evaluation of a rule involves a pair of nested if statements:

```

    if(precondition)
      if(condition)
        return score_one
      else
        return score_two
    end if
  else
    return 0
  end if
  
```

4. APRIORI ASSOCIATION RULE ALGORITHM

Association rule learning in data mining is a technique for discovering interesting relations between variables in large databases.

An example rule for the supermarket could be “{butter, bread} -> milk” meaning that if butter and bread is bought, customers also buy milk.

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 $X \rightarrow Y$ is the percentage of transaction in the database that contains XUY. Confidence for the association rule is $X \rightarrow Y$ is the ratio of the number of transaction that contains XUY to the number of transaction that contain X [6]. The flowchart for Apriori Association Rule algorithm is shown in figure 2.

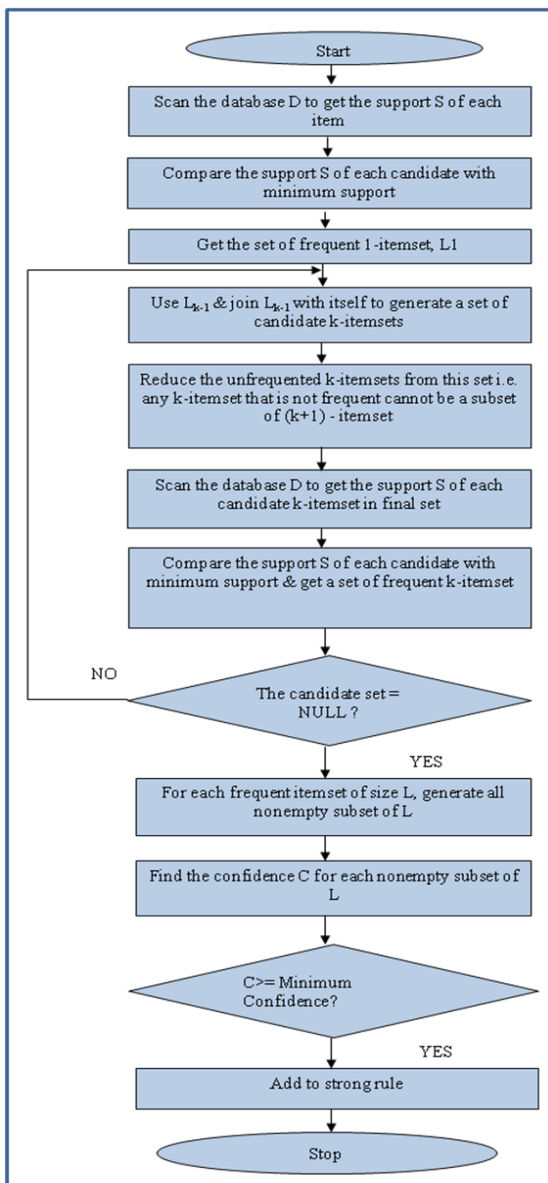


Figure 2: Flowchart for Apriori Association Rule Algorithm

5. RESULT & IMPLEMENTATION

Here we are considering the sample data extracted from Moodle database of a college after collection of data for course enrollment by student as shown in Table 1. In this table, we consider 45 student & 15 courses. Fifteen courses are 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 particular course & no represent that student do not like that course. In preprocessing step, we prepare the table after deleting 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 rows i.e. 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 result of preprocessing of data is shown in table 2.

The graph for sample table before preprocessing & after preprocessing is shown in figure 3 & 4 respectively. The number inside the circle represents the course count & right side of circle represents the course name. Figure 5 represents the graph for sample table after application ADTree classification algorithm.

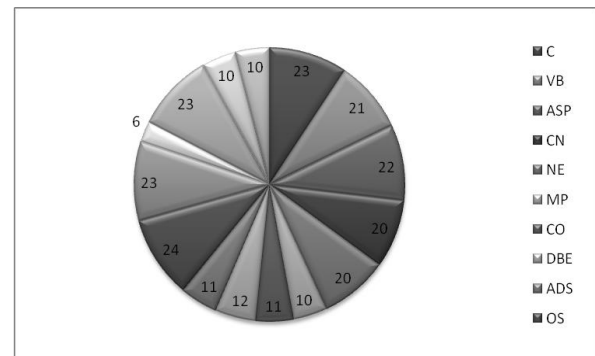


Figure 3: Graph for Sample table before preprocessing

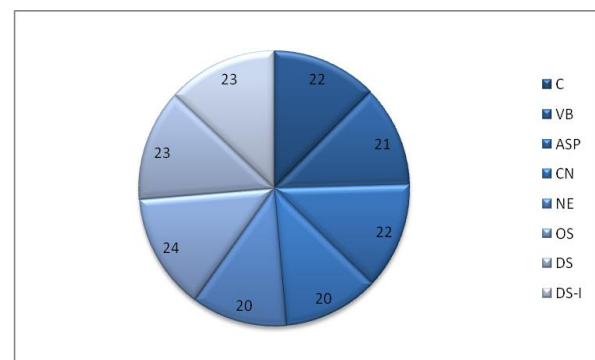


Figure 4: Graph for Sample table after preprocessing

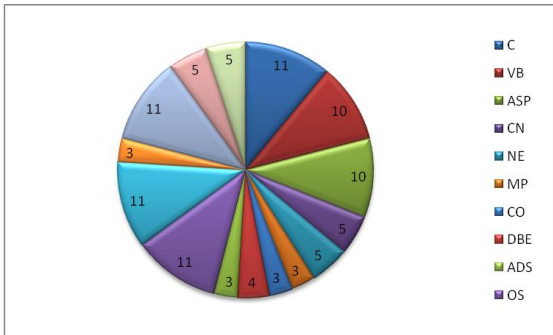


Figure 5: Graph for sample table after application of ADTree classification algorithm

To test the result we use the open source data mining tool Weka. 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 [6]. The result of applying Apriori association rule before & after preprocessing of data

is shown in first & second row of table 4. 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 4. 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. If we increase the support then we get the refined rule which is shown in second row of table 4.

For combination of classification & association rule, there is no need to preprocess the data. First we apply the ADTree classification algorithm & then the Apriori association rule algorithm. After classification of data, we got the table 3. The result of applying the Apriori Association Rule algorithm on classified data shown in table 3 is shown in third row of table 4.

Table 1: Sample data from Moodle database [7]

Courses→ Roll_No↓	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 [7]

Subject → Roll No of Student ↓	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: Table after application of classification algorithm-ADTree

Courses → Roll_No ↓	C	VB	ASP	CN	NE	MP	CO	DBE	ADS	OS	DS	FSA	DS-I	SE	STQA
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: Result after application of machine learning algorithms

Course considered	Parameter 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. Computer_Organization=no → Microprocessor=no 2. Database_Engg=no → Advanced_Database_System=no 3. Computer_Organization =no Finite_State_Automata=no → Microprocessor=no 4. Microprocessor =no → Computer_Organization =no 5. Software_Testing_& Quality_Assurance=no → Software_Engg=no 6. Software_Engg =no → Software_Testing_& Quality_Assurance =no 7. Advanced_Database_System =no → Database_Engg =no 8. Microprocessor =no Finite_State_Automata =no → Computer_Organization =no 9. Finite_State_Automata =no Software_Testing_& Quality_Assurance =no → Software_Engg =no 10. Finite_State_Automata =no Software_Engg =no → Software_Testing_& Quality_Assurance =no

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. Distributed_System=yes → Operating_System=yes 2. Visual_Basic=yes → Active_Server_Pages=yes 3. Network_Engg=yes → Computer_Network=yes 4. Computer_Network =yes → Network_Engg =yes 5. C_Programming=yes Visual_Basic =yes → Active_Server_Pages =yes 6. Distributed_System =yes Data_Sstructure_I=yes → Operating_System =yes 7. Operating_System =yes → Distributed_System =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
	Minimum support: 0.6 Minimum metric <confidence>: 0.9	Best rules found: 1. Distributed_System =yes → Operating_System =yes 2. Operating_System =yes → Distributed_System =yes
After Application of Classification algorithm-ADTree & 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	Best rules found: 1. Operating_System =yes → C_Programming =yes 2. C_Programming =yes → Operating_System =yes 3. Distributed_System =yes → C_Programming =yes 4. C_Programming =yes → Distributed_System =yes 5. Data_Sstructure_I =yes → C_Programming =yes 6. C_Programming =yes → Data_Sstructure_I =yes 7. Distributed_System =yes → Operating_System =yes 8. Operating_System =yes → Distributed_System =yes 9. Data_Sstructure_I =yes → Operating_System =yes 10. Operating_System =yes → Data_Sstructure_I =yes

6. CONCLUSION AND FUTURE WORK

As ADTree classification algorithm & Apriori association rule algorithm are the popular technique of data mining, we try to use the combination of these two algorithms to find the best combination of courses in educational system. The advantage of applying this combination is that there is no need to preprocess the data as we have to preprocess if we are using only the association rule mining. First we apply the ADTree classification algorithm to the sample data obtained from Moodle course of a college & then apply the Apriori Association Rule algorithm to the classified data. We compare the result after application of combination of these two algorithm & result after application of Apriori Association Rule mining. We found that as we increase the support, we get the refined rule using Apriori Association Rule but we get the less number of rules as well as we have to preprocess the data to use this algorithm. If we use the combination algorithm then we get the association rules which match the real world interdependencies among the courses. Future work includes the atomization of this combination algorithm.

7. REFERENCES

- [1] Al'ipio M. Jorge, Paulo J. Azevedo: "An experiment with association rules and classification: post-bagging and conviction?" Supported by the POSI/SRI/39630/2001/Class Project
- [2] Ye-Zheng Liu, Yuan-Chun Jiang , Xiao Liu , Shan-Lin Yang : "CSMC: A combination strategy for multi-class classification based on multiple association rules" Knowledge-Based Systems 21 (2008) 786–793
- [3] P. R. Pal, R.C. Jain: "Combinatorial Approach of Associative Classification" International; Journal Advanced Networking and Applications
- [4] B. Ramasubbarreddy, A. Govardhan & A. Ramamohanreddy: "Classification Based on Positive and Negative Association Rules" International Journal of Data Engineering, (IJDE), Volume (2): Issue (2) : 2011 84
- [5] 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
- [6] 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)

- [7] 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**
- [8] Han,J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd edition.
- [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-12, proceeding published by IEEE
- [10] Fernando Berzal, Juan-Carlos Cubero, Nicolás Marín Daniel Sánchez, Jose-María Serrano, Amparo Vila."Association rule evaluation for classification purposes" Actas del III Taller Nacional de Minería de Datos y Aprendizaje, TAMIDA2005, pp.135-144 ISBN: 84-9732-449-8 © 2005 Los autores, Thomson
- [11] "Alaa Al Deen" Mustafa Nofal and Sulieman Bani-Ahmad: "Classification Based On Association-Rule Mining Techniques: A General Survey And Empirical Comparative Evaluation" accessed from <http://www.scribd.com/doc/46190273/Classification-Based-on-Association-rule-Mining-Techniques-a-General-Survey-and-Empirical-Comparative-Evaluation-Ubiquitous-Computing-and-Communicat> on 14-02-2012
- [12] Alternating decision tree, available at: http://en.wikipedia.org/wiki/alternating_decision_tree. Accessed on 13-02-2012.