Asian Journal of Computer Science and Technology
ISSN: 2249-0701 Vol.7 No.3, 2018, pp. 72-75

Finding Relationship among Various Students' Attributes with Sport-Interest using Association Rule Mining Method

Sanjeev Gour
Department of Computer Science, Career College, Bhopal, Madhya Pradesh, India
E-Mail: sunj129@gmail.com

***Abstract*-Data mining methods are widely used in educational domain for the purpose of finding useful information from the large student's database. This information is then used to understand the behaviors of studentsin respect of their academic and other curricular performance. One of such Data mining methods called Association rule mining is used in this research study to analyze the student's database of Career College Bhopal using two mining tools called Weka and XLMiner.The database contain records of 212 students with main attributes like Student's Gender, Category, Subject, name of district where he/she belong and their parent's/guardian's occupation/profession and sport-interest. Sport department of any educational institute also need to understand the behavior and psychology of students for their sport interest to make sport-policy for their institute. In this paper, author has found some unknown relationship among these attributes with respect to sport-interest which is a target attribute. This experimental study has generated many association rules that can be used to answer the questions like which student from particular course, district and category will participate in sports? Which sport usually prefer by male and female student most? Which student can be performed better in which sport? In this way sport policy maker can use these mined information about the sport interest of students to make better decisions in sport framework in an educational institute.**

***Keywords*: Association Rule Mining, Educational Data Mining, Apriori Algorithm, Weka, XLMiner**

I. INTRODUCTION

Today, huge amount of transactional data are generating everyday in every organization. This data can play an important role to make a better decision policy in any organization by extracting hidden patterns from them. Data Mining is nowadays an emerging and powerful technology used in almost every application domain to discover hidden knowledge from the existing secondary database [1]. Many educational institutes now implementing data mining techniques to find the useful knowledge from their databases, this is known as (EDM) Educational Data Mining [2]. As today every educational instititute generate large amount of data of their students in every years, various EDM techniques are used to analyze these data items which can help to make better decision policy in teaching, learning and other curricular activities ([3], [4], and[5]).

Sport is also one of main curriculum activity in an educational institute. Sport department of any educational institute also need to understand the behavior and psychology of students for their sport interest to make sport-policy for their institute. Sport policy provides a framework to support physical education and sport in school and colleges. Data mining tools can also be very useful to this application domain. If the institute has sufficient data about their student's interest in sport activity, then these data can be analyzed to extract useful information. This information can be used to answer the questions like which student from particular course, district, category will participate in sports?, which sport usually prefer by male and female student most? Which student can be performed better in which sport? In this way sport policy maker can use these mined information about the sport interest of students to make better decisions in sport framework in an educational institute.

II. METHODOLOGY

A. Association Rule Mining

To extract or discover hidden relationship or patterns among data items of the large database, Association rule mining is one of most common and popular method of Data mining.It contains some powerful algorithms which have great capability to find hidden patterns, relationships or correlations among set of data items of largedatabase. This method is most suitable in such applications domain where they have large data set. It is capable to find useful and interesting association rules which can be used to make right decision for a policy in any organization. It generates simple If => Then rules which can be easily interpreted. If => Then rule usually have following structure[6]:

If Antecedent (A) =>then Consequent (B) [support, confidence].

Where Support indicates the frequency of the rule in the transactions of database. If it's value is high it means the rule contain a great part of database. Confidence shows the degree at which each rule antecedent implies the rule consequent. In this way high confidence show strong association and high support indicates how strongly their association are. Another important measure is Lift Value which treats as ratio of the confidence of the rule and the expected confidence of the rule.

For this study Apriori algorithm has been selected for finding frequent data item sets in this transaction student's

database.The Apriori algorithm is most common and important algorithm for finding hidden relationship among dataset as it is so simple to learn and use[7].

III. DATA MINING TOOLS

There are many online software tools free available nowadays for analyzing large data. For this study, author use two data mining tools called Weka and XLMiner.

Weka isan emerging and powerful data mining tool which contains number of recent machine learning algorithms and visualization features. All data mining tasks such as data cleaning, preprocessing, clustering, classifications, associations etc can easily be implemented in this tool. It is very user friendly and easy to operate as all the methods can be run by only single click. This is why; this tool is used to analyze the students' dataset.

Another tool called XLMiner is used for finding association rules from the same dataset.It is one of online analyzing tool which includes various statistical and data mining techniques. All the features of this cloud based tool is incorporate with in-house Excel software so implementing all necessary features are very easy and user friendly. Like Weka, this tool has also all capabilities to generate, save and visualize a statistical or data mining model. This tool is also used to discover hidden relationship among various attributes of student's data with their sport-interest field.

IV. DATABASE

The Student's Databaseof department of Computer Science was collected from the Admission Cell of Career College Bhopal for this study. The databasecontain some of main attributes/fields of student like Student's Gender, Category, Subject, Name of district where he/she belong and their parent's/guardian's occupation/profession and sport-interest. The database contains the student's records which were admitted or enrolled in 2016-17 of Computer department of Career College. This dataset has been generated in excel format by college's local admission software at the time of admission. For the study author has selected six main attributes/fields (Table I) of student dataset and generate an association model to find the relationship among these attributes with sport-interest field. The data of more than 200students were in Excel format. The Original format of dataset with selected fields has been shown in Table I.

TABLE I SAMPLE STUDENTS 'DATA BASE TABLE WITH SELECTED ATTRIBUTES

S. No.	Gender	Subject	Category	District	Father-Profession	Sport-Interest
1	Male	Computer Science	GEN	Bhopal	Government Employee	Yes
2	Female	Computer Science	GEN	Chhatarpur	Government Employee	No
3	Male	BCA	OBC	Hoshangabad	Business	No
4	Female	M.Sc.	GEN	Bhopal	Business	No
5	Male	BCA	OBC	Bhopal	Farmer	Yes

V. EXPERIMENTAL PROCESS

The experimental process to find the association rules using both mining tools are briefly discussed here.

A. WEKA Environment

1. Loading Data- Load the dataset in Weka explorer by importing data file
2. Preprocessing-Transform the format of input data intoformat compatible with tools. i.e. Save the .xls or .csv as .arff file format. Also choose the filter command for data filtering and cleaning.
3. Set the type and role of each attributes in dataset.i.e. Set the support and confidence minimum values
4. Select Apriori Algorithm and Run the Associate Command.
5. Interpret the rules generated in If=>Then format.

B. XLMiner Environment

1. Loading Data-Load the dataset in XLMiner by clicking Get Data tab.
2. Preprocessing-Transform the format of input data into format compatible with tools. i.e. Transform the .xls file into .csv file format.
3. Set Sampling and Fitting parameters .i.e. Method, minimum value for support and confidence etc.
4. Select Apriori Algorithm and Run the Associate Command.
5. Interpret the rules generated in If=>Then format.

VI. RESULTS AND DISCUSSION

Many association rules are generated by Weka and XLMiner tool (Figure 1 and 2). Some of strongly associated and interesting rules are presented here:

The percentage ratio measured for choosing the attribute sport-interest (Yes: No) by the students of department of computers science is 56:44 respectively. The analysis clearly show that sport-interest attribute has selected approximately in equal ratio(5:4) by the students but the ratio measured for choosing the same attribute by male and female students is approximately 8:2 respectively. This analysis clearly shows that girls student dropping out of

sports at four times the rate of boys from department of computer science.

One of Association Rule generated by Apriori method in Weka with highest lift value (1.63) indicates (Figure-1) that students of Electronic course are showing more interest in participating in sport activity than other Courses. Another Rule generated by XLMiner (Figure-2) support this result with lift value 3.78 that male students from electronic branch strongly opted for sport-interest attribute/field.

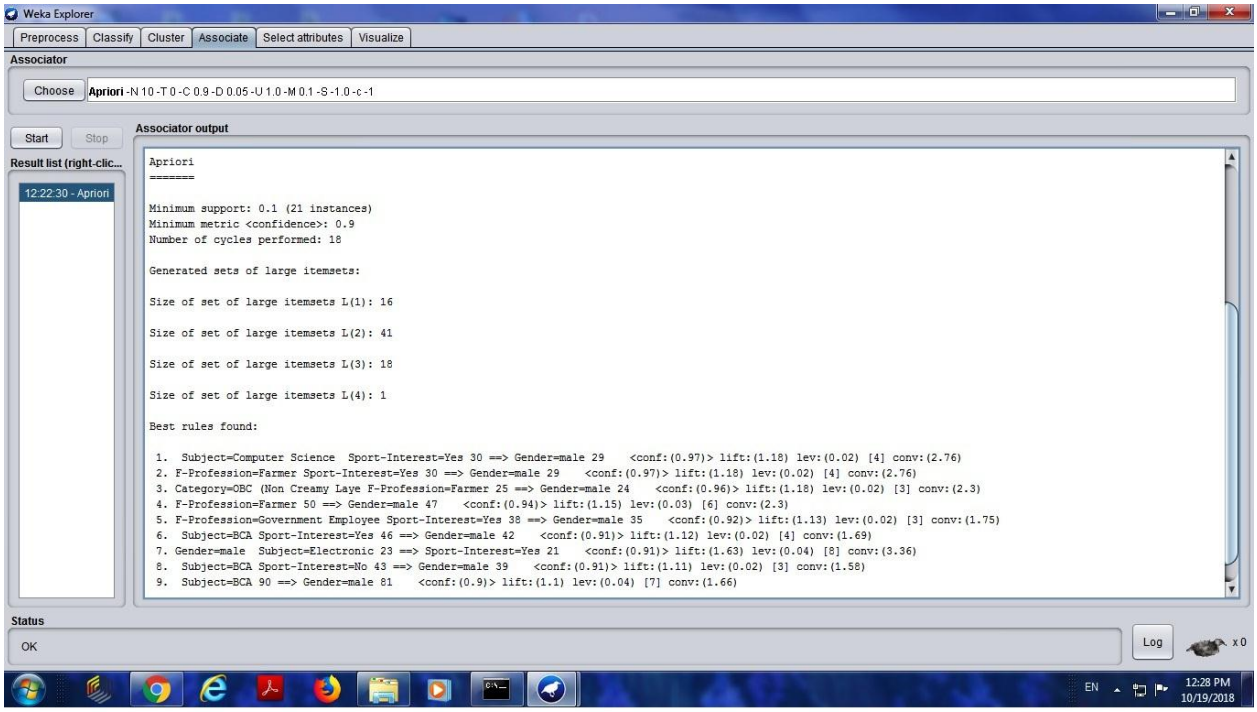

Fig. 1 Screenshot of Result Window in Wekafor Association Rule

Rule ID	A-Support	C-Support	Support	Confidence	Lift-Ratio	Antecedent	Consequent
Rule 1	27	52	20	74.07407407	0.897435897	[GEN]	[male]
Rule 2	11	27	7	63.63636364	1.484848485	[female]	[GEN]
Rule 3	27	28	17	62.96296296	1.416666667	[GEN]	[No]
Rule 4	28	27	17	60.71428571	1.416666667	[No]	[GEN]
Rule 5	13	27	8	61.53846154	1.435897436	[Business]	[GEN]
Rule 6	10	27	6	60	1.4	[Information Technology]	[GEN]
Rule 7	12	52	10	83.33333333	1.009615385	[Computer Science]	[male]
Rule 8	21	52	19	90.47619048	1.096153846	[OBC (Non Creamy Laye]	[male]
Rule 9	52	35	32	61.53846154	1.107692308	[male]	[Yes]
Rule 10	35	52	32	91.42857143	1.107692308	[Yes]	[male]
Rule 11	19	52	16	84.21052632	1.020242915	[Government Employee]	[male]
Rule 12	10	52	9	90	1.090384615	[Electronic]	[male]
Rule 13	17	52	15	88.23529412	1.069004525	[Farmer]	[male]
Rule 14	27	52	24	88.88888889	1.076923077	[BCA]	[male]
Rule 15	28	52	20	71.42857143	0.865384615	[No]	[male]
Rule 16	9	52	8	88.88888889	1.076923077	[SC]	[male]
Rule 17	19	52	14	73.68421053	0.892712551	[Bhopal]	[male]
Rule 18	13	52	10	76.92307692	0.931952663	[Business]	[male]
Rule 19	10	52	8	80	0.969230769	[Information Technology]	[male]
Rule 20	7	52	7	100	1.211538462	[Hoshangabad]	[male]
Rule 21	7	52	6	85.71428571	1.038461538	[Private Job]	[male]
Rule 22	11	52	10	90.90909091	1.101398601	[Other]	[male]
Rule 23	12	35	9	75	1.35	[Computer Science]	[Yes]
Rule 24	21	35	15	71.42857143	1.285714286	[OBC (Non Creamy Laye]	[Yes]
Rule 25	17	21	10	58.82352941	1.764705882	[Farmer]	OBC (Non Creamy Laye]
Rule 26	19	35	13	68.42105263	1.231578947	[Government Employee]	[Yes]
Rule 27	10	35	9	90	1.62	[Electronic]	[Yes]
Rule 28	17	35	9	52.94117647	0.952941176	[Farmer]	[Yes]
Rule 29	27	35	14	51.85185185	0.933333333	[BCA]	[Yes]
Rule 30	9	35	6	66.66666667	1.2	[SC]	[Yes]

Fig. 2 Screenshot of Result Window in XLMiner for Association Rule

Another interesting and strong association found that students from the BCA coursewhich were belongs to Science subjects did not opted the sport-interest attribute/field while students from the BCA course which were belongs to Commerce, Arts or other subjects are opted for sport-interest attribute/field.

Some of rules generatedby XLMiner and Weka tool are showing an interesting association that students whose parents are Government employee have selected the sport-interest attribute/field with Liftand Confidence value more than 1.5 and 92 % respectively.

Some of rules mined by XLMiner show that students whose parents are Farmers and belongs to OBC category also opted for sport-interest attribute with lift value 2.0.

The XLMiner generate a rule which indicate with 100% confidence that male students of computer science belonging to Hoshangabad district picked the sport-interest attribute.

Many such useful and interesting association rules are generated. These rules can be interpreted in terms of associations or correlations among various data items (attributes) of student with their sport interest attribute.

VII. CONCLUSION

This study is of course an application of educational data mining in that author has used well known Apriori algorithm of Association rule mining to analyze the student's database of Career College Bhopal. The database contains much information about their choices which they opted at the time of admission. Sport-interest is one of attribute/choice, they have been selected. This study, finds many useful relationship among various attributes of student with their sport interest. These extracted associations can be used to understand the behavior of student towards his/her sport interest. Forthis study, author has used two most commonly data mining tool called Weka and XLMiner. This study can be extended to finding associations among more attributes of student database like NSS/NCC activity, Training and Placement, Student Council and other curricular activities.

REFERENCES

[1] Sunita B. Aher , L.M.R.J. Lobo, "Data Mining in Educational System using WEKA", *International Conference on Emerging Technology Trends (ICETT)*, Proceedings published by International Journal of Computer Applications, pp. 20-25,2011.

[2] Elakia, Gayathri and Aarthi, J. Naren, "Application of Data Mining in Educational Database for Predicting Behavioral Patterns of the Students", *International Journal of Computer Science and Information Technologies*, Vol. 5, No. 3, pp. 4649-4652, 2014.

[3] AbdulmohsenAlgarni, "Data Mining in Education", *International Journal of Advanced Computer Science and Applications*, Vol. 7, No. 6, pp. 456-461, 2016.

[4] Pooja M .Dhekankar , Dinesh S. Datar ,"Analysis of Student Performance by using Data Mining Concept", *International Journal on Recent and Innovation Trends in Computing and Communication*, Vol. 3, No. 5, pp. 2942-2944, 2015.

[5] DorinaKabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classification", *Cybernetics And Information Technologies*, Vol. 13, No. 1, pp. 61-72, 2013.

[6] MamtaGour, SanjeevGour, Purushottam Sharma, "Developing a Water Quality Model for Upper Lake of Bhopal Using Data Mining Methods",*International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, Vol. 4, No. 12, pp. 4320-4324, 2015.

[7] ShaileshJaloree, SanjeevGour and MamtaGour, "Water Quality Assessment using Association Rule Mining for River Narmada",*Indian Journal of Science and Technology*, Vol. 9, No. 10, pp. 1-5, 2016.