Usage of Apriori Algorithm of Data Mining as an Application to Grievous Crimes against Women

Divya Bansal[#1], Lekha Bhambhu[*2]

[1]M.Tech(CSE & GJU, Hisar), JCDM Engineering College,Sirsa Haryana(INDIA)
[2] Assistant Professor (CSE), JCDM Engineering College,Sirsa Haryana(INDIA)

***Abstract*- Quantitative data must be converted into qualitative data, for this association algorithm only can apply to it. As association rule deals with frequent item sets as done by many association algorithms such as: Apriori algorithm, that's why in most real life applications Apriori algorithm is used. In this paper author contains the use of association rule mining in extracting patterns that occur frequently within a dataset and showcases the implementation of the Apriori algorithm in mining association rules from a dataset which is manual collection of demeaning crimes against women which is collected from Session court. In this paper author considers the two Association Rule algorithms i.e. Apriori Algorithm and PredictiveApriori Algorithm and compares the result of both the algorithms using WEKA, a data mining tool.As result of rules of both algorithms clearly shows that Apriori algorithm performs better and faster than PredictiveApriori algorithm.**

***Keywords*-Data Mining, Association Rule, Apriori Algorithm, Command line interface.**

I. INTRODUCTION

Data Mining is a detailed process of analyzing large amounts of data and picking out the relevant information. It refers to extracting or mining knowledge from large amounts of data. The data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically. [4,13]. Association Rule in Data Mining plays a important role in the process of mining data for frequent itemsets. Finding frequent patterns called associations. Frequent patterns are the patterns that occur frequently in the data. Patterns can include itemsets, sequences and subsequences. A frequent itemset refers to a set of items that often appear together in a transactional data set. example : bread and milk. .It involves the following steps: cleaning and integrating data from data sources like databases, flatfiles, pre-treatment of selecting and transformation target data, mining the required knowledge and finally evaluation and presentation of knowledge. A data mining algorithm is complete if it mines all interesting patterns. It is often unrealistic and inefficient for data mining systems to generate all possible patterns. Instead, user-provided constraints and interestingness measures should be used to focus the search . In data mining, association rule learning is a most popular methodology to identify the interesting relations between the data stored in large database.

II. RELATED DEFINATION

Association Rule: Association rule of data mining involves picking out the unknown inter-dependence of the data and finding out the rules between those items [3]. Agrawal introduced association rules for point of sale (POS) systems in supermarkets. A rule is defined as an implication of the form A=>B, where A∩ B≠Ø. The left-hand side of the rule is called as antecedent. The right-hand side of the rule is called as consequent.

Support: I = { i1,i2,i3, … , im} is a collection of items. T be a collection of transactions associated with the items. Every transaction has an identifier TID [6]. Association rule A=>B is such that A Є I, B Є I. A is called as Premise and B is called as Conclusion. The support ,S, is defined as the proportion of transactions in the data set which contains the itemset. Support(X=>Y) = Support (XUY) = P(XUY).

Confidence: The confidence is defined as a conditional probability Confidence (X=>Y) = Support (XUY) / Support(X) = P(Y/X).

Lift: is the ratio of the probability that L and R occur together to the multiple of the two individual probabilities for L and R, i.e.,

$$\text{lift} = \Pr(L,R) / \Pr(L).\Pr(R).$$

Conviction: is similar to lift, but it measures the effect of the right-hand-side not being true. It also inverts the ratio. So, a conviction is measured as:

$$\text{conviction} = \Pr(L).\Pr(\text{not } R) / \Pr(L,R)$$

III. APRIORI ALGORITHM

A realization of frequent pattern matching based on support and confidence measures produced excellent results in various fields. As Table1 gives the Psedocode of apriori algorithm.

TABLE I APRIORI ALGORITHM

Join Step: C_k is generated by joining Lk-1with itself

Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset

C_k: Candidate itemset of size k
L_k: frequent itemset of size k
L_1= {frequent items};
for(k= 1; L_k!=∅; k++) do begin
C_{k+1}= candidates generated from L_k;
for eachtransaction tin database do
increment the count of all candidates in C_{k+1}that are contained in t
L_{k+1}= candidates in $C_{k+1}$with min_support
end
return∪$_k$$L_k$;

A. Working of Apriori Algorithm:

In general, Apriori Algorithm can be viewed as a two-step process:
(i) Generating all item sets having support factor greater than or equal to, the user specified minimum support.

(ii) Generating all rules having the confidence factor greater than or equal to the user specified minimum confidence [8].
Example:
A database has five transactions. Let the min sup = 50% and min con f = 80%. As it shows the transaction in Figure1

Step 1:Find all Frequent Itemsets, as shown in Figure 2
Frequent Itemsets:
{A},{B},{C},{E},{A,C},{B,C},{B,E},{C,E},{B,C,E}
Step 2: Generate strong association rules from the frequent itemsets. Results are shown in Table 2

TID	ITEMS
100	A C D
200	B C E
300	A B C E
400	B E

Fig 1 Database

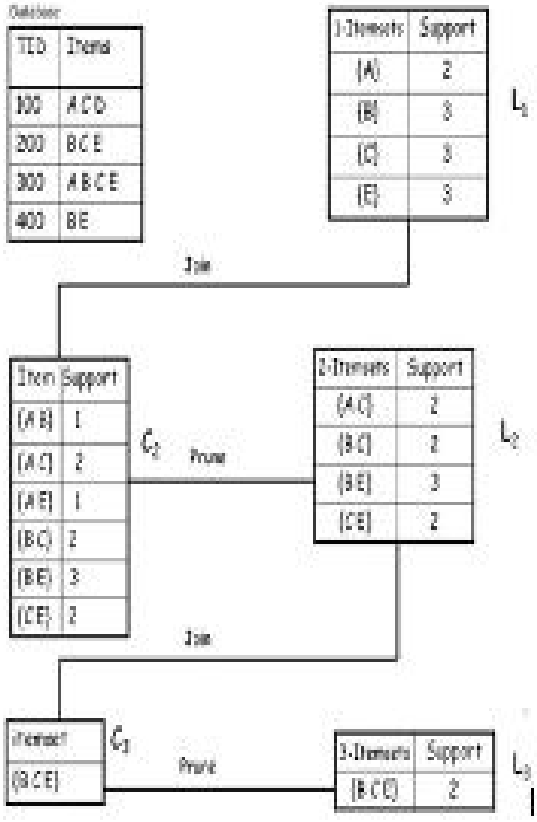

Fig 2 frequent itemsets

TABLE II SUPPORT & CONFIDENCE

Rules	Support(XY)	Support(X)	Confidence
{A}->{C}	2	2	100
{B}->{C}	2	3	66.66
{B}->{E}	3	3	100
{C}->{E}	2	3	66.66
{B}->{C E}	2	3	66.66
{C}->{B E}	2	3	66.66
{E}->{B C}	2	3	66.66
{C}->{A}	2	3	66.66
{C}->{B}	2	3	66.66
{E}->{B}	3	3	100
{E}->{C}	2	3	66.66
{C E}->{B}	2	2	100
{B E}->{C}	2	3	66.66
{B C}->{E}	2	2	100

As it incudes all the frequent itemsets.

IV. IMPLEMENTATION OF APRIORI ALGORITHM

In the implementation of the Apriori algorithm in mining association rules from a dataset containing cases of different crimes against women as dataset available in Session court. Extraction of frequent item sets is essential towards mining useful and relevant patterns from datasets. As it includes data under section 376,363,366. As it includes data of various section which comes under the Pathetic crimes against Women, as data is collected from Session court Sirsa, and Rewari .

A. WORKING OF WEKA: As it includes attributes such as Age of boy,Age of Girl, Relation ,Section. As Relation Attribute tells us about the what a relation a victim has with a accused.WEKA is used to figure out all this. Figure 3 shows importing of database to WEKA.[19]

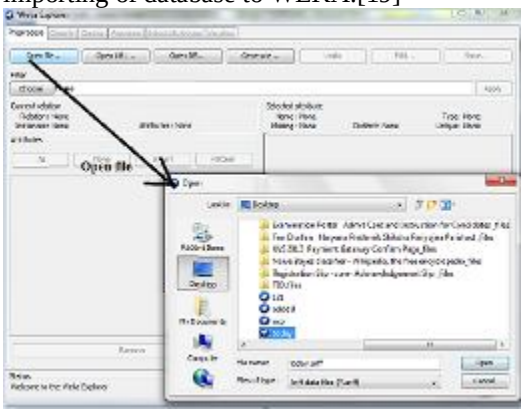

Fig 3 to Import database

1) *Preprocess Panel:* The preprocess panel is the start point for knowledge exploration. From this panel you can load datasets, browse the characteristics of attributes. Figure 4 shows the preprocess panel of womencrime dataset.

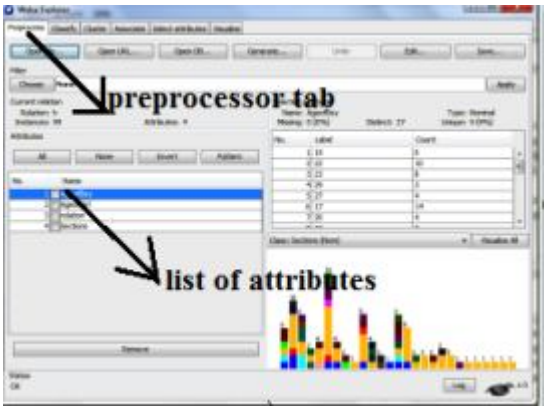

Fig 4 Preprocessor

2) *Associate Panel:* From the associate panel you can mine the current dataset for association rules using the weka associators. Different options available for Apriori are class index, lower bound, min support, metric type, minimum metric, number of rules etc shown in Figure 5, Figure 6 shows the ten best association rules using Apriori.

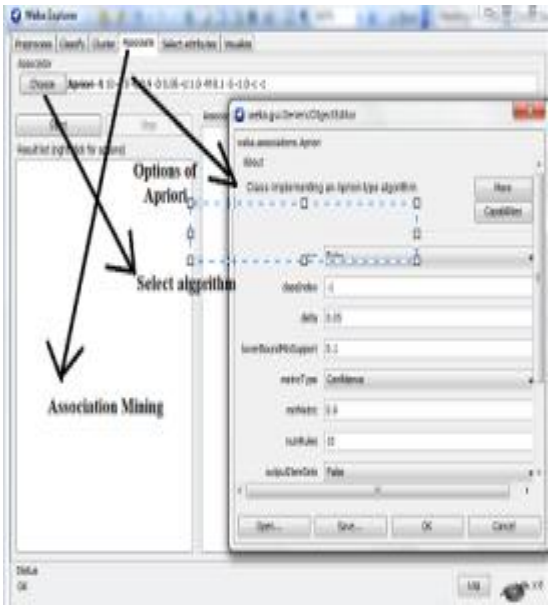

Fig 5 Selecting parameters

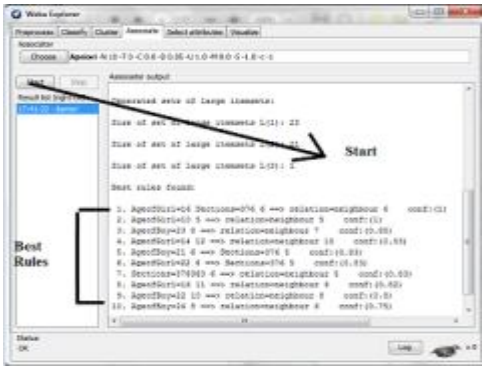

Fig 6 Best Rules

3) *Command Line Interface:* (CLI) is used shown in Figure 7. type command in space given below. Figure 8 shows the association rules and frequent itemsets for Apriori using CLI.

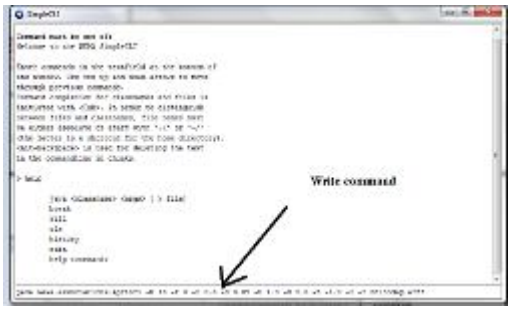

Fig 7 CLI

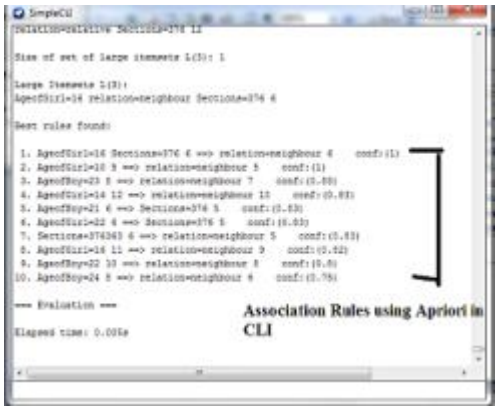

Fig 8 Association rules in CLI

V. EXPERIMENTAL RESULTS

Here the experimental results of both algorithms of Association Rule Mining are given. In this paper comparison has done on both Apriori algorithm and PredictiveApriori algorithm. As Apriori is explained in previous section, now the brief discussion on other algorithm. As elapsed time is calculated for both the algorithms of association with the help of Command line interface (CLI) of WEKA.

A. PredictiveApriori Association Rule Mining:

In predictive Apriori association rule algorithm, support & confidence is combined into a single measure called "Accuracy□. {Support, Confidence}=> Accuracy. In this predictiveApriori association rule algorithm, this predictive accuracy is used to generate the Apriori association rule. In Weka, this algorithm generates "n" best association rule based on "n□ is number of rules specified by the user.

B. Comparative Results:

This paper finds the result using Association rule algorithms by mining tool WEKA. In this women's crimes
Dataset is used for comparison with 4 attributes and 99 instances. Table 3 represents the result of Apriori Association rule Algorithm and Table 4 represents the results of PredictiveApriori algorithm. Figure 9 shows the time comparison of both the algorithms.
This paper clearly shows that age group of male is 20-25 who are doing this heartmelting crimes aginst girl are generally the known to girl and they lived near by girl house and age group of girls i.e. 16-21 and in this paper it is Apriori algorithm is more faster than PredictiveApriori algorithm.

TABLE III APRIORI ALGORITHM

Apriori Association Rule Algorithm	
List of Attributes	Best Rules Found
1.Age of boy 2.Age of girl 3.Relation 4.Section	1. AgeofGirl=16 Sections=376 6 ==> relation=neighbour 6 conf:(1)) 2. AgeofGirl=10 5 ==> relation=neighbour 5 conf:(1) 3. AgeofBoy=23 8 ==> relation=neighbour 7 conf:(0.88) 4. AgeofGirl=14 12 ==> relation=neighbour 10 conf:(0.83) 5. AgeofBoy=21 6 ==> Sections=376 5 conf:(0.83) 6. AgeofGirl=22 6 ==> Sections=376 5 conf:(0.83) 7. Sections=376363 6 ==> relation=neighbour 5 conf:(0.83) 8. AgeofGirl=16 11 ==> relation=neighbour 9 conf:(0.82) 9. AgeofBoy=22 10 ==> relation=neighbour 8 conf:(0.8) 10. AgeofBoy=24 8 ==> relation=neighbour 6 conf:(0.7
Elapsed Time of Apriori Association rule Algorithm 0.072s	

TABLE IV PREDICTIVEAPRIORI ALGORITHM

PredictiveApriori Association Rule Algorithm	
List of Attributes	Best Rules Found
1.Age of boy 2.Age of girl 3 Relation 4 Section	1. AgeofGirl=16 Sections=376 6 ==> relation=neighbour 6 acc:(0.98057) 2. AgeofGirl=10 5 ==> relation=neighbour 5 acc:(0.97436) 3. AgeofGirl=18 3 ==> relation=neighbour 3 acc:(0.94532) 4. AgeofGirl=21 3 ==> relation=relative 3 acc:(0.94532) 5. AgeofBoy=40 2 ==> relation=neighbour 2 acc:(0.90973) 6. AgeofBoy=45 2 ==> Sections=376 2 acc:(0.90973) 7. AgeofGirl=26 2 ==> Sections=376 2 acc:(0.90973) 8. AgeofGirl=4 2 ==> relation=neighbour 2 acc:(0.90973) 9. AgeofGirl=24 2 ==> relation=relative 2 acc:(0.90973) 10. AgeofBoy=23 8 ==> relation=neighbour 7 acc:(0.82238)
Elapsed Time of PredictiveApriori Association rule Algorithm- 0.724s	

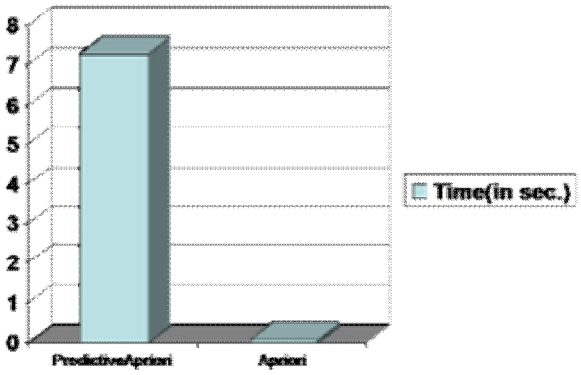

Fig 9 Comparison of Predictiveapriori & Apriori

VI. CONCLUSIONS AND FUTURE WORK

The purpose of research is to discover answer to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered yet. Through each research study has its own specific purposes. Apriori Algorithm is used to discover and understand the underlying patterns involved in the court's records from their data contains in various sections. Molestation has become an alarming public issue not only in one or the other area but of world wide issue. Hence, there is a need present for accurate, timely information to react to changing pathetic condition of women, identifying who are mostly involved i.e. age group of accused, stranger or known to the victim, and basically which age groups girls are the main target of victims are analyzed to improve the deteriorating condition of women. As this research works answers all the questions as agr gropu pf men is 20- 24 ,age group of girls who are on their target is 16-22 and mostly accused are well known by the victim. This is helpful for the government ,society and police that they will take certain actions towards the male society. It basically tells what steps a society should take so that this appalling situation of women will improved and women can go freely without *feardom but with freedom.*

Future work: includes the detection of fraud cases, many of cases are untrue. To collect data on child abused cases and tell what people are involved in such crimes.

REFRENCES

[1] Association Rule Learning – Wikipedia, the free encyclopedia.

[2] Chen, H., Zhan, Y., Li, Y., (2010), "The application of Decision Tree in Chinese email Classification". In the proceeding of 9th International Conference on machine Learning and Cybernetics, 2010, pp. 305-308.

[3] Feng Yucai, "Association Rules Incremental Updating Algorithm", Journal of Software, Sept., 1998.

[4] Jaiwei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Second Edition, Morgan Kaufmann Publishers.

[5] Lee, Y. D., Tasi, W. C., Hsu, S.F., (2009), "Exploring the Relationship between Parental Information Literacy and Regulating Rules in Family by Data Mining". In the proceeding of International Conference on Computational intelligence and security, 2009, pp. 21-24.

[6] Lei Guoping, Dai Minlu, Tan Zefu and Wang Yan, " The Research of CMMB Wireless Network Analysis Based on Data Mining Association Rules", IEEE conference paper – project supported by the Science and Technology Research Project of Chongqing municipal education commision under contract no KJ101114 and KJ 111103, 2011.

[7] Lin, H., Goumin ,Z., Liu, Q., (2009), "Application of Apriori Algorithm to Data Mining of the Wildfire". In the proceeding of 6th International Conference on Fuzzy Systems and Knowledge Discovery, 2009, pp. 426-429.

[8] Maragatham G and Lakshmi M(2012) " A RECENT REVIEW ON ASSOCIATION RULE MINING" Maragatham G et al./ Indian Journal of Computer Science and ISSN : 0976-5166E) Vol. 2 No. 6 Dec 2011-Jan 2012.

[9] Merseron, A. and Yacef, K., "Interestingness Measures for Association Rules in Educational Data". In the proceeding of 1st International Conference on Educational Data Mining, 2008, pp. 1-10.

[10] Patil, B.M., Toshniwal, D., Joshi, R.C.,(2012)," Analytical Study Using Data Mining for Periodical Medical Examination of Employees." Volume 174, 2012, pp 221-227

[11] Prasad, P., Malik, L., (2011), "Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions". In the International Journal on Computer Science and Engineering, 2011, pp. 2177-2182.

[12] Rakesh Agrawal, Tomasz Imielinski and Arun Swami (1993)," Database Mining: A Performance Perspective" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL 5, NO. 6. DECEMBER 1993

[13] S.Suriya,Shantharajah and Deepalakshmi (2012)" A Complete Survey on Association Rule Mining with Relevance to Different Domain" INTERNATIONAL JOURNAL OF ADVANCED SCIENTIFIC AND TECHNICAL RESEARCH, ISSN:2249- 9954ISSUE2VOLUME 1 (FEBRUARY 2012).

[14] www.codeproject.com/Articles/70371/Apriori-Algorithm/Omar Gamil

[15] Yang, G., Zhero, H., Wang, L., Liu,Y., (2009), "An implementation of improved Apriori Algorithm". In the proceeding of 8th International Conference on Machine Learning and Cybernetics, 2009, pp. 1565-1569.

[16] Yashoda, P., Kannan, M., (2011), "Analysis of a Population of Diabetic Patients Databases in Weka tool". In the International Journal of Scientific and Engineering Research Volume 2, 2011, pp. 1-5.

[17] Zawaidah, Jbara and Zanona (2011) " An Improved Algorithm for Mining Association Rules in Large Databases" World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 1, No. 7, 311-316, 2011.

[18] Zhu, Z., Wang, J., (2007), "Book recommendation on service by improved association rule mining algorithms ". In the proceeding of international conference on Machine Learning and Cybernetics, 2007, pp. 3864-3869.

[19] Ms.Shweta and Dr. Kanwal Garg(2013)," Mining Efficient Association Rules Through Apriori Algorithm Using Attributes and Comparative Analysis of Various Association Rule Algorithms" International Journal of Advanced Research in Computer Science and Software Engineering ISSN: 2277 128X Volume 3, Issue 6, June 2013.

[20] Savasre, A, Omienciski,E, and NavatheS,(1995),"An efficient algorithm for mining association rules in large databases"In the proceeding of 21st international conference on VLDB,1995,pp 432-444.

[21] Tan, P., Steinbach, M., Kumar, V., (2006)," Introduction to Data Mining". Addison Wesley,2006.

[22] Sumithra, R., Paul, S., (2010), "Using distributed apriori association rule and classical apriori mining algorithms for grid based knowledge discovery". In the proceeding of 2nd International Conference on

Computing, Communication and Networking Technology, 2010, pp. 1-5.

[23] Thabtah, F., Cowling, P., Hammoud, S., (2005),"Improving rule sorting, predictive accuracy and training time in associative classification". An article in Press of Expert Systems with Applications, (2005), pp. 1–13

[24] http://www.newsagepublishers.com/researchmethodology.

[25] http://www.wikipedia.org/wiki/research.

[26] B.M.Patil,Durga,Toshniwal, R. C. Joshi,(2009)."Predicting Burn Patient Survivability Using Decision Tree In WEKA Environment."IEEE International Advance Computing Conference (IACC 2009) Patiala, India, 6-7 March 2009.