

EVOLVING ASSOCIATION RULE USING WEKA FOR THE EXTRACTION OF PATTERNS IN ONLINE TRANSACTIONS

Shikha Dubey (MCA)*

Dr. Shivaji D. Mundhe**

ABSTRACT:

Data mining is used for mining data from databases and finding out meaningful patterns from the database. Association rule mining is an important component of data mining. Association rules are an important class of methods of finding patterns in data. Various Open source tools are available for experimentation. In this paper we have tried to evolve association rule using Weka and based on it a new algorithm will be developed. Many organizations are now using these data mining techniques. Weka formally called Waikato environment for knowledge learning is an open source which supports many different data mining tasks such as data preprocessing, classification, clustering, association and feature selection.

Keywords: *Association Rules, Rule Induction Technique, Apriori Algorithm, WEKA.*

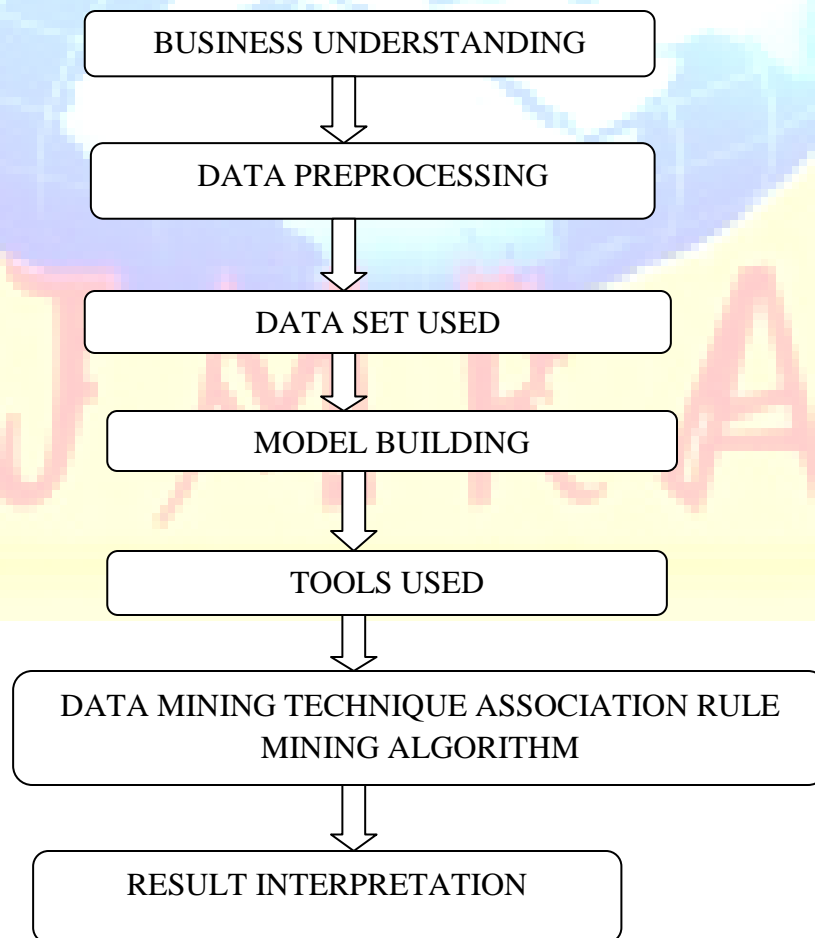
* *Assistant Prof. D. Y. P. I. M. R., Pimpri, Pune-18*

** *Director SIMCA, Pune-41*

Introduction:

The proposed methodology focuses on e business activities and using a multilayer data mining approach, provides high level of business intelligence for enterprises that utilize ecommerce as their main transactional model. To improve the performance of business there is a need to design and implement an effective algorithm based on association rule and therefore the researcher has attempted to work on it.

State-of-the-art open-source data mining suites of today have come a long way from where they were only a decade ago. They offer nice graphical interfaces, focus on usability and interactivity, support extensibility through augmentation of the source code or (better) through use of interfaces for add-on modules Some of the open source software that the researcher has studied are R, Tanagra, Weka, YALE, KNIME, Orange, GGobi, KEEL..

Methodology Being Carried Out For Study:

Objectives:

- i) To illustrate the use of WEKA for association rule mining.

- ii) To implement the algorithm and test it for real time datasets to reduce costly, repeated database scans

It is presumed that the result of the proposed research work will help the industry for finding the accurate and useful mining information related to their data and also help in research of similar aspects and for anyone who likely to develop the interest in similar type of work. The proposed work will also help to reduce the present problem related to mining such as the repetitive I/O disk scan, huge computation involved during the candidate generation and high memory dependency. So this work is proposed to reduce costly, repeated database scans as well as CPU time without compromising on accuracy.

Methodology:

Weka formally called Waikato environment for knowledge learning supports many different data mining tasks such as data preprocessing, classification, clustering, association and feature selection. It is open source software originally written in C and completely rewritten in java. Consider the transactions of an online Shoppe which offers items like {shirt, trouser, belt, tie, wallet, shoes, socks} for sale purpose. For better marketing it is always better for the store to arrange the items in a format in which maximum sales occur.

So in this study a sample of transactions are taken to find out the association among various items. In Weka the main screen appears after that we have to load our CSV sheet or .arff format sheet. Value of 1 indicates that item is bought in the transaction and 0 indicate that item is not bought in the transaction.

Figure I: WEKA GUI Chooser

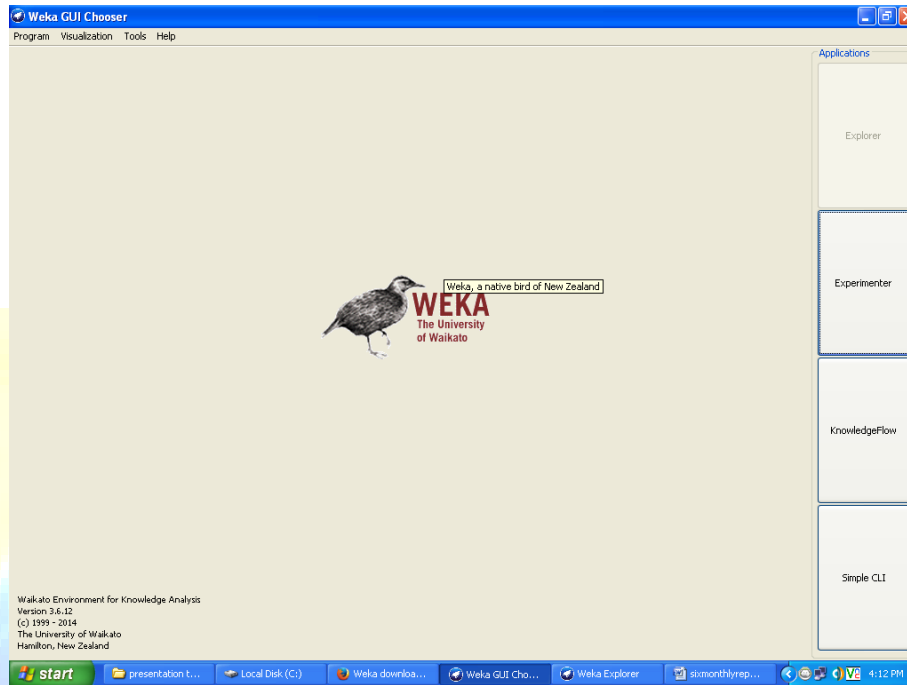


Figure II: Sample Dataset

Figure III: Open & Load CSV file

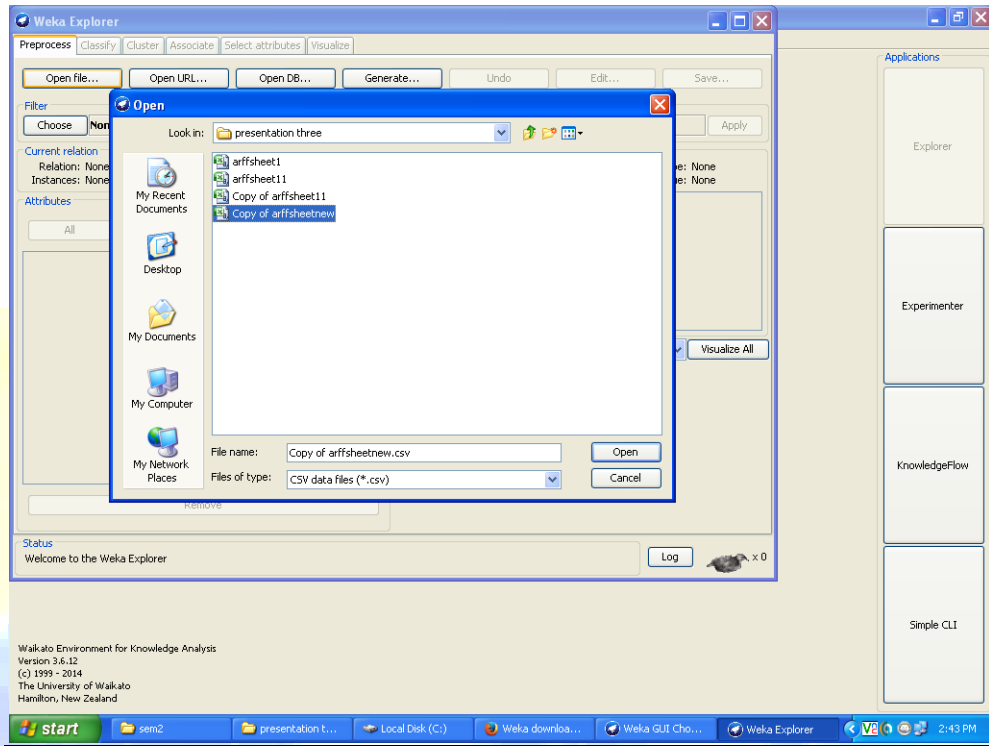
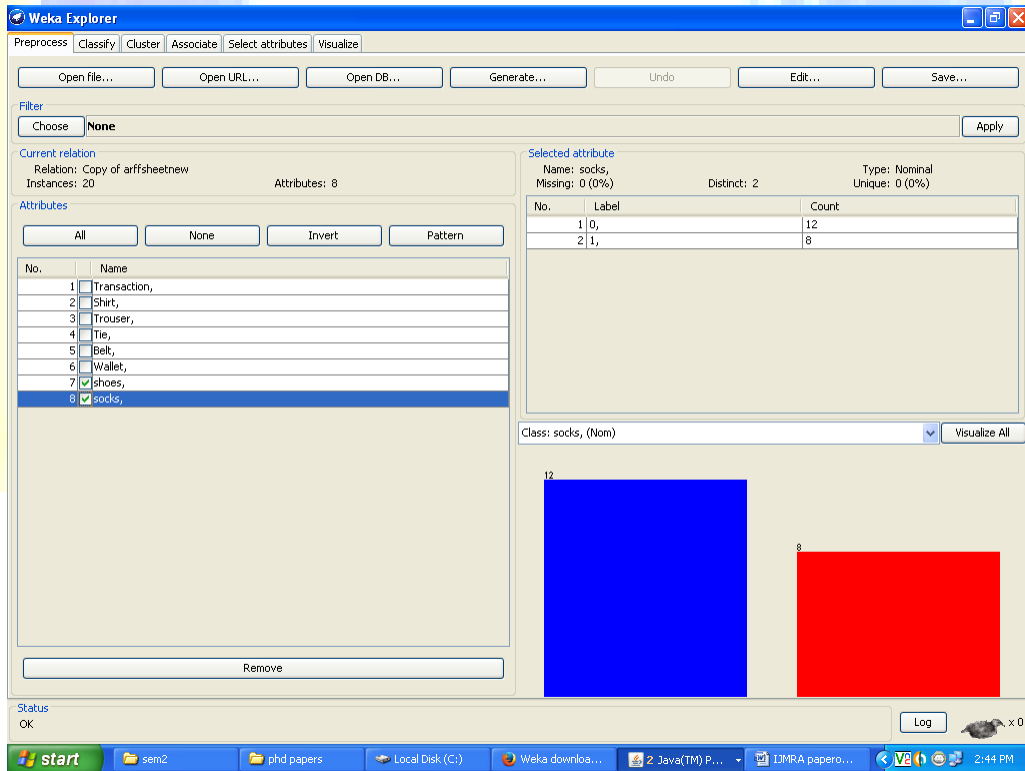
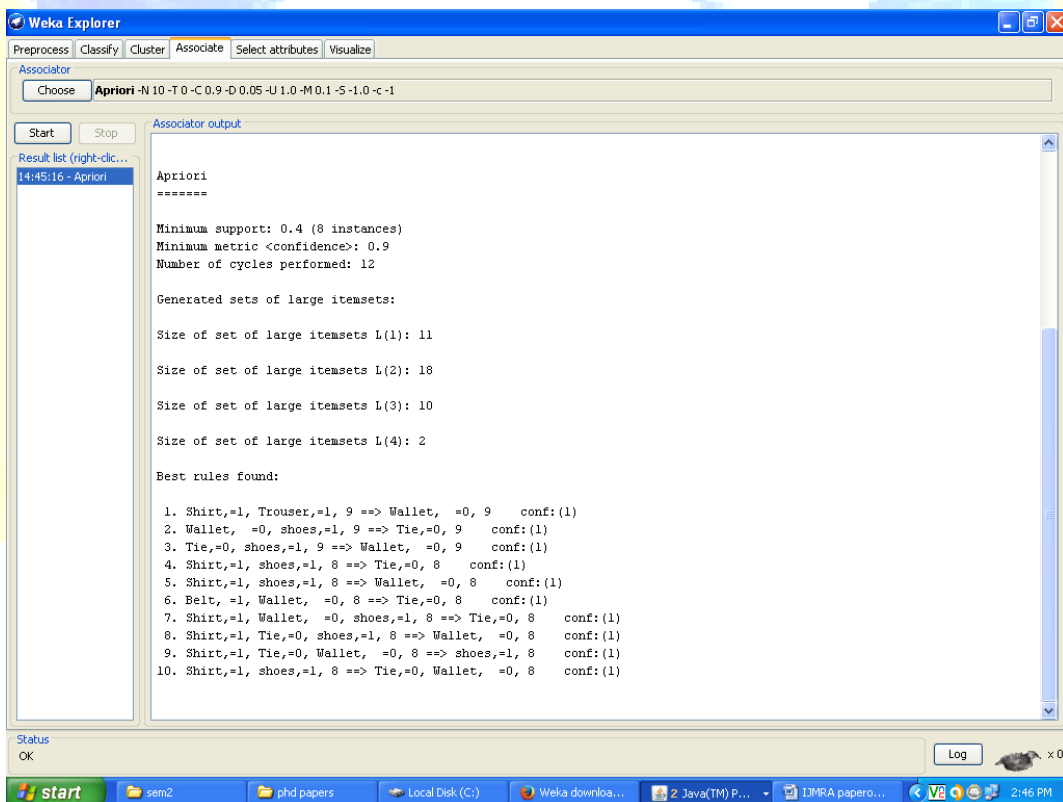
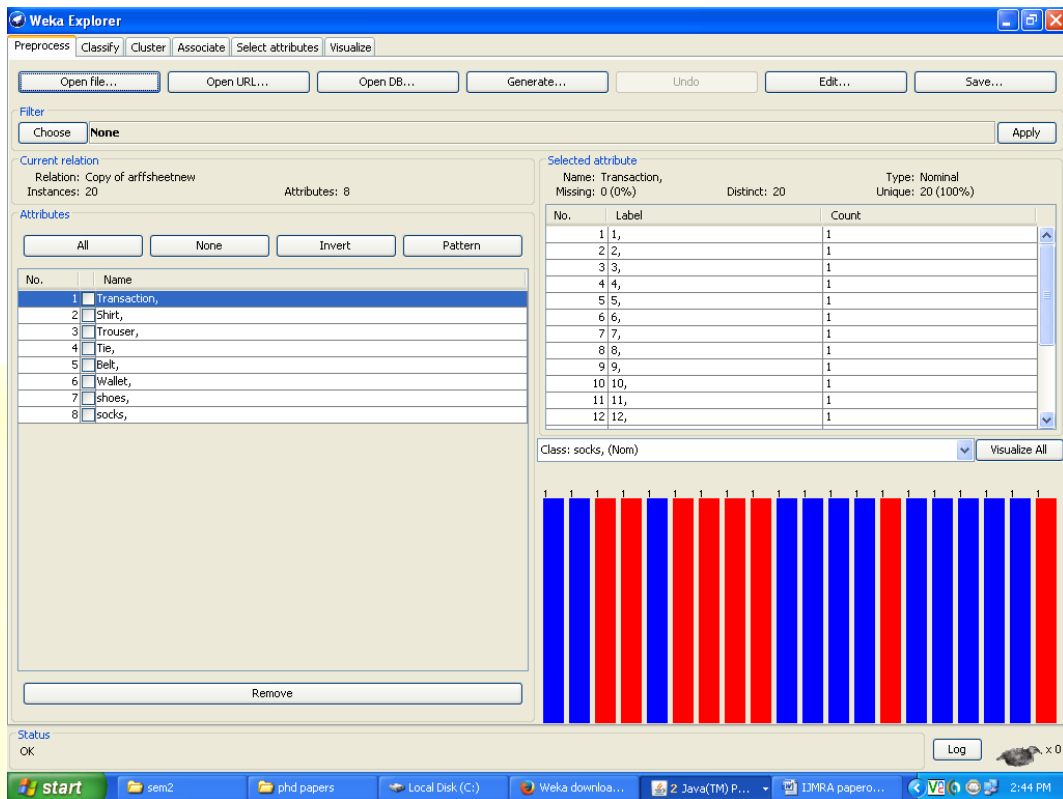
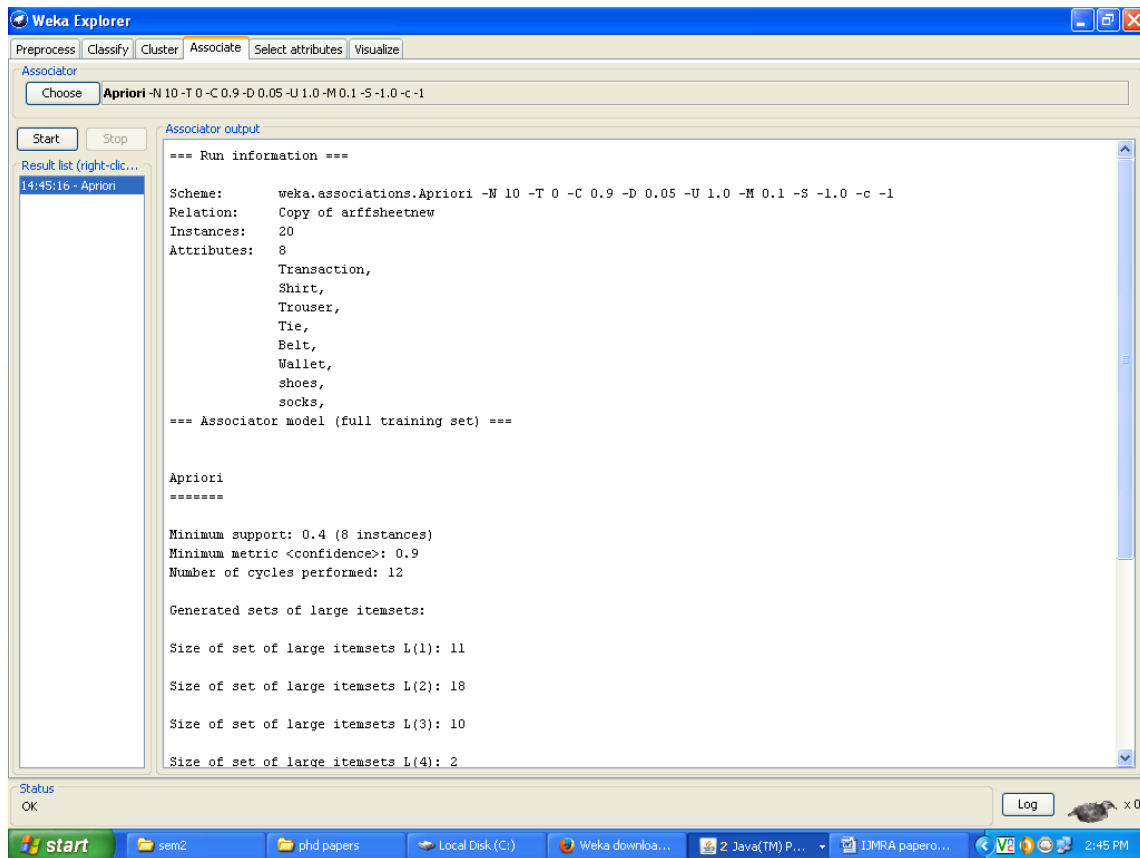


Figure IV: Association rules generated thru Weka







Conclusion:

Through this study it was found that shirt and trouser have a greater association having a higher level of confidence along with shoes and socks, based on this the placement of items in a website can be decided i.e. the items which have higher association should be kept nearer to increase the sales. For my experimentation I tested the working of algorithm in WEKA. Further it will be implemented in C#.Net Application. An algorithm which will reduce the Number of passes of Apriori will also be developed.

REFERENCES:

- 1/ Bing Liu¹, Wynne Hsu², Yiming Ma Integrating Classification and Association Rule Mining, Appeared in *KDD-98*, New York, Aug 27-31, 1998
- 2/ D.Narmadha¹, G.NaveenSundar², S.Geetha³ An Efficient Approach to Prune Mined Association Rules in Large Databases: *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814, 2011
- 3/ Farah Hanna AL-Zawaidah¹, Yosef Hasan Jbara², 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.*

4] Masoud Pesaran Behbahani¹, Islam Choudhury², Souheil Khaddaj³: Enhancing Organizational Performance through a New Proactive Multilayer Data Mining Methodology: An Ecommerce Case Study *International Journal of Innovation, Management and Technology*, Vol. 3, No. 5, October 2012

5] Mohd Helmy Abd Wahab¹, Mohd Norzali Haji Mohd², Mohamad Farhan Mohamad Mohsin³: Discovering Web Server Logs Patterns Using Generalized Association Rules Algorithm, Intech ISBN: 978-953-307-067-4, 2010

6] Qiankun Zhao¹, Sourav S², Bhowmick³: Association Rule Mining: A Survey, Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116, 2003

7] Rakesh Agrawal¹, Ramakrishnan Srikant²: Fast Algorithms for Mining Association Rules, IBM Almaden Research Center
650 Harry Road, San Jose, CA 95120, 2009

8] Ulrich Guntzer¹, Gholamreza Nakhaeizadeh²: Algorithms for Association Rule Mining – A General Survey and Comparison SIGKDD Explorations Volume 2 issue 1, 2000.

9] Pramod Prasad et al. using association rule mining for extracting product sales patterns in retail store transaction *International journal of computer science and engineering* Vol.3 may 2011.

10] Zahir Tari¹, Wensheng Wu²: ARM: a hybrid association rule mining algorithm Springer journal, 2006

