

# PERFORMANCE IMPROVEMENT USING INTEGRATION OF ASSOCIATION RULE MINING AND CLASSIFICATION TECHNIQUES

Dr. V. Vaithiyathan<sup>1</sup> K. Rajeswari<sup>2</sup> Rahul Pitale<sup>3</sup> Kapil Tajane<sup>4</sup>

**Abstract**— Data mining is an emerging field to find interesting patterns from a large collection of data sets. Classification algorithms like neural network, decision trees are used to classify the patterns according to class labels in the input patterns. Association rule mining is used for finding related or more frequent patterns found in a given data set. This paper integrates both these techniques, namely classification and association. Experiments are carried out on different classification techniques and association techniques using Weka. The main aim of this paper is to find accuracy using different classification methods and association rule mining technique. The paper proposes a new methodology of integrating both association rule mining and classification and investigated their performance and compared their results. IRIS data set from University of California, Irvine is used for evaluation purpose.

**Index Terms**— Association, Classification, DataMining, IRIS, J48, MultilayerPerception, Weka.

## 1 INTRODUCTION

DATA Mining is a technique which process large amount of data. It is one of the technique which helps to focus on information present on their large databases[5]. There are many techniques and tools available in data mining which discovers the patterns and relationship among large data sets. Classification and association are the popular techniques used to predict user interest and relationship between those data items which has been used by users. Classification methods includes Bayesian network, J48 Decision tree, Neural Network etc. Association is used to find all co-occurrence relationship. It has some famous applications like Market Basket Data Analysis. There are many algorithms used in association rule mining like Apriori algorithm it generate all frequent itemsets and it also generate support and confidence count for data sets.

Our next section presents literature survey on Classification rule mining, Association Rule Mining and their algorithms. In Section 3 Proposed methodology along with Results and Discussion is given. In Section 4 a brief idea about Weka tool is given. Finally the last section is for conclusion.

## 2 LITERATURE RIEW

**Classification:** Classification Rule Mining aims to discover a small set of rules in a database to form an accurate classifier[5]. Classification has many methods such as Decision tree, neural network, etc, Mainly there are two steps to implement classification functions first in build classification model to describing a predetermined set of classes or concepts. Second step, Classification. There are many methods used in classification[7][8],e Bayes Network, J48 Decision tree, etc.

### 2.1 Bayes Network:

It is based on bayes theorem and it uses probability theory for classification. In this model they gives random variables and conditional probability using acyclic graph representation[7]. This bayes network are used in IR, image processing, computer game application etc. For conditional Independencies bayes network work in natural way, it is easy for implementation. It convert conditional independencies between graphical format. It is popular method because it handles uncertainty in simple way. It uses network so it is different from other probability approaches.

- 
- Dr.V.Vaithiyathan, Associate Dean Research, CTS Chair Professor, SASTRA University, Tanjore, India. E-mail: [vvv@it\\_sastra.edu](mailto:vvv@it_sastra.edu).
  - K Rajeswari is currently pursuing PhD in SASTRA University, India, PH-9226094965. E-mail: [raji.pccoe@gmail.com](mailto:raji.pccoe@gmail.com).
  - Rahul Pitale is currently pursuing masters degree in Computer Engineering in Pune University, India, PH-9420240035. E-mail: [rahulpitale3@gmail.com](mailto:rahulpitale3@gmail.com).
  - Kapil Tajane is currently pursuing masters degree in Computer Engineering in Pune University, India, PH-9823400036. E-mail: [kapiltajane@gmail.com](mailto:kapiltajane@gmail.com).

### 2.2 J48 Decision Tree:

Decision tree is one of the important methods in classifier. It developed decision tree based on input data [7]. J48 algorithm used in weka. It is one of the efficient way to represent the information and it gives better performance in structured data. Pruning is done in j48 it produces results in very easy manner. This pruning is done until all leaf is pure. This process gives maximum accuracy. It also known as statistical classifier. It is based on C4.5 algorithm.

### 2.3 Multilayer Perception:

An artificial neural network model that maps sets of input data onto a set of correct results. Directed graph uses multiple layers to depict multilayer perception. It uses backpropagation technique is modification of the standard linear perception in that it uses three or more layers of neurons powerful than perception in that it can distinguish data that is not inearly seperable, or sepearable by hyper-palne. The Multilayer perception consist of input and an output layer with one or more hidden layers of nonlinearly activating nodes Each node is one layer connects with a certain weight  $W_{ij}$  to every node in the following layer. This Multilayer perception is used in speech recognition, and machine learning software, also used on cyber security.

### Association Rule Mining:

Associaition means how the attributes are closely engaged with one another [4]. The problem of association rule mining can be stated as follows let  $A = \{A_1, A_2, \dots, A_n\}$  be a set of items [4]. let  $P = (p_1, p_2, \dots, p_n)$  be a set of transaction, where each transaction  $t_i$  is a set of items such that  $p_i \subseteq A$ . An association rule is an implication of the form  $x \rightarrow y$  where  $x \subseteq A$   $y \subseteq A$  and  $x \cap y = \phi$

$x$  (or  $y$ ) is a set of items, called an itemset. A transaction  $p_i \in P$  is said to contain an item set  $x$ . If  $x$  is a  $\subseteq p_i$  (we also say that the itemset  $x$  covers  $p_i$ ) support count of  $x$  in  $p$  is the number of transactions in  $P$  that contains  $x$ . The strength of the rule is measured by it's support and confidence. Apriori algorithm is used in association rule mining algorithms. It works in two steps it generate all frequent itemsets and also generate all confident association rules from frequent itemsets. The apriori algorithms relies on apriori or downward closure property to generate all frequent itemsets there are some examples in association rule mining are finding patterns in biological data bases, extraction of knowledge from software Engg metric, Web Mining, text mining etc.

## 3 PROPOSED WORK:

In this paper, Classification and Association methods, Using Weka we take one datasets from UCI the data set is IRIS, It

contains 5 attributes sepalength, sepalwidth, petalength, petalwidth, and class. we perform some data mining operations on IRIS data sets, like preprocessing, classification and association and clustering. our first approach to find accuracy of IRIS data sets so we classify this datasets using some Classification methods like BayesNet, J48, MultilayerPerception etc. The results are given as in Table 3.1.

Table 3.1 Accuracy for different classifiers

| Name of the Classifier | Accuracy % |
|------------------------|------------|
| BayesNet               | 92.67%     |
| Multilayer Perception  | 97.33%     |
| J48                    | 96%        |

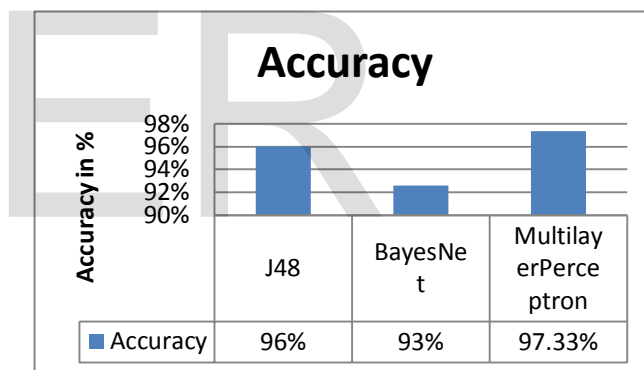


Fig.3.1 Compariosn of Accuracy

Using discretization and Association Rule Mining we perform the same on classification methods. Better Accuracy is obtained in Bayes Function combined with association. Association gives the most frequent attributes Table

Table 3.2 Accuracy for different classifiers

| Name of the Classifier | Accuracy % |
|------------------------|------------|
| BayesNet               | 94%        |
| Multilayer Perception  | 90.67%     |
| J48                    | 96%        |

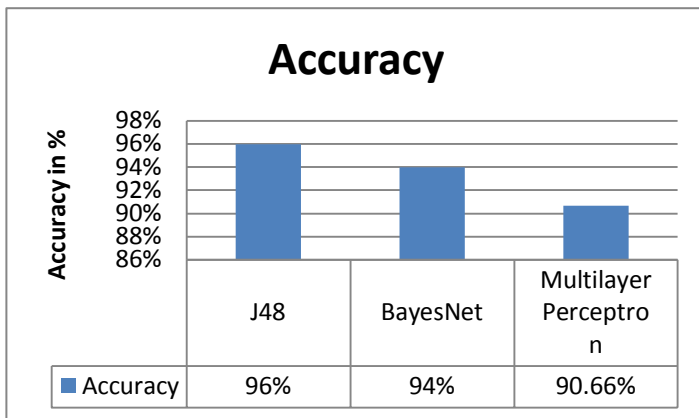
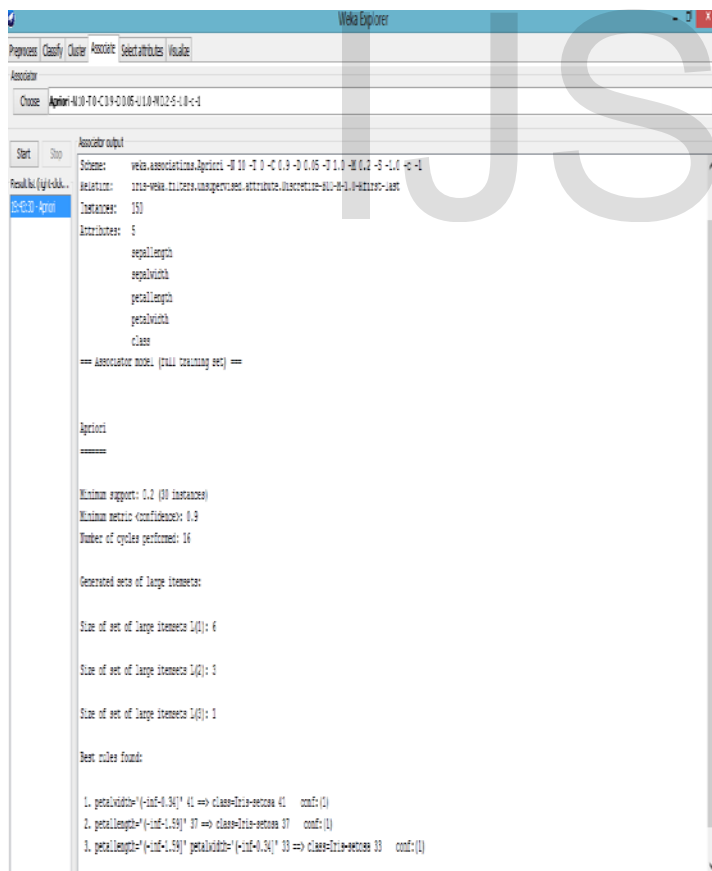
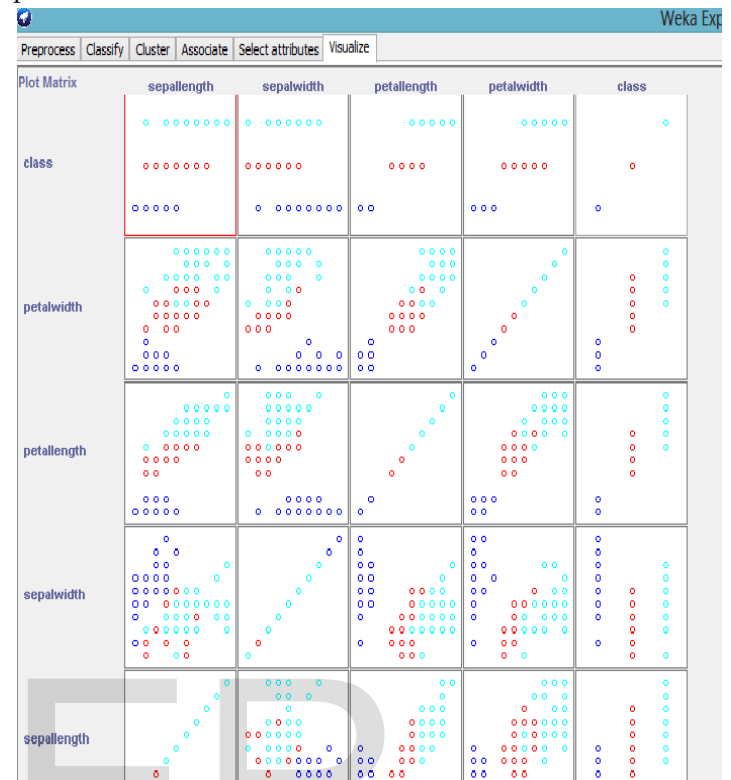


Fig.3.2 Comparison of Accuracy

In our proposed work neural network has one drawback i.e it takes much amount of times than J48 and BayesNet so our proposed our will not feasible for neural network implementation. In our proposed work we use association rule mining in that we use apriori algorithms for generating rules from IRIS datasets , In this we method we can increase the minimum support count from 0.1 to 0.2 we can get the most frequent attributes values this results are shown below.



From Above results we can say that using Combination of both Association and Classification we can get the better output. Some cluster also form to show the results is as follows.



#### 4 RESOURCES USED FOR PROPOSED WORK:

**Weka:** It is abbreviation of Waikato Environment for knowledge Analysis is popular suite of machine learning software written in java[2], developed at the Waikato university. Weka Supports several Standards data mining tasks, i.e preprocessing, clustering, classification, regression, and visualization and feature selection. UCI: UCI machine learning repository is a collection of databases[3], It is developed in 1987 since that time it has used by student, researcher, and educators.

#### 5 CONCLUSION AND FUTURE SCOPE:

Classification aims to discover a small set of rules in databases to form an accurate classifier and association finds all rules in databases that satisfy some minimum support and minimum confidence. A new methodology to find more accurate results using an Integration of classification and association rule mining is proposed here. Experimentation is performed using weka. Our results show that using Bayes method of classification, accuracy is improved after finding associated items. In future our work will focus performance issues in Neural network and their accuracy with a combination of clustering

with classification.

## REFERENCES:

- [1]Purohit, S.; Kulkarni, A., "Credit evaluation model of loan proposals for Indian Banks," Information and Communication Technologies (WICT), 2011 World Congress on , vol., no., pp.868,873, 11-14 Dec. 2011 doi: 10.1109/WICT.2011.6141362
- [2]Machine Learning Group at the University of Waikato <http://www.cs.waikato.ac.nz/ml/weka/>
- [3]Frank, A. & Asuncion, A. (2010). "UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science.
- [4]A. M. Khattak, A. M. Khan, Sungyoung Lee, and Young-Koo Lee, "Analyzing Association Rule Mining and Clustering on Sales DayData with XLMiner and Weka", Vol. 3, No. 1, March, 2010
- [5]Neelamadhab Padhy, Dr. Pragnyaban Mishra and Rasmita Panigrahi," The Survey of Data Mining Applications And Feature Scope",International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT), Vol.2, No.3, June 2012
- [6]Ian Harris, Jorg Denzinger and Dean Yergens ,"Application of the Weka Machine Learning Library to Hospital Ward Occupancy Problems" , December 12, 2007
- [7]Mohd Fauzi bin Othman,Thomas Moh Shan Yau, "Comparison of Different Classification TechniquesUsing WEKA for Breast Cancer".

## AUTHORS BIOGRAPHY



**Dr. V. Vaithyanathan** is a Professor Associate Dean Research, CTS Chair Professor, SASTRA University. His area of interest are data mining and Image Processing. He is associated with many funded projects. He has published many International papers in reputed journals and conferences



**K.Rajeswari** received BE and Mtech in Computer Science

.She has published several papers in Data Mining. She is winner of Cambridge International Certification for Teachers with Distinction. She has about 15 years of Teaching experience Currently she is working as an Associate Professor Computer Science and Engg in PCCOE,Akurdi,Pune.She is pursuing her Ph.D in SASTRA University,Tanjore,India.



**Rahul Pitale** received Bachelor degree in Computer Science & Engg from Amravati University. Currently he is pursuing Master of Computer Engg. From PCCOE Pune University



**Kapil Tajane** received Bachelor degree in Computer Science & Engg from Amravati University. Currently he is pursuing Master of Computer Engg. from PCCOE, Pune University