Identification of Frequent Item Search Patterns Using APRIORI Algorithm and WEKA Tool

K.RAGHAVENDRA SWAMY
Associate professor, Department of CSE
RISE Krishan Sai Group of Institutions
Ongole, Prakasam Dist., A.P., India

G.HARINADH BABU
Associate professor, Department of CSE
RISE Krishan Sai Group of Institutions
Ongole, Prakasam Dist., A.P., India

R.VENKATASUBBAIAH,
Associate professor, Department of CSE
RISE Krishan Sai Group of Institutions
Ongole, Prakasam Dist., A.P., India

Abstract: Identification and Analysis of frequent item search patterns will play a key role in Business data processing and Data Analytics. We have so many Data Analytics techniques to analyze Business Data. In this study paper we summarizes various Data Mining techniques on frequent item search patterns using Apriori Algorithm and Weka Tool, Associative classification, Minimum support confidence, candidate Generations and visualizations are used to analyze frequent item search patterns.

Keywords: Frequent Item Search Patterns, Weka Tool, Visualization, Data Mining, Association Rule Mining, Frequent Item Set and Pruning.

I. INTRODUCTION

Data mining is an essential step in the process of knowledge discovery in databases in which Intelligent methods are applied in order to extract patterns. In this paper we identify and analyze the Frequent item patterns in Market basket of business processing. we know that Market basket analysis is used for to identify the most associated items and this type of analysis used for the Business improvement and introduction of new products and study the different transactions, how the different transactions are carried out and how the business needs are changed. In this paper we are using Apriori algorithm to generate new List items and Candidate keys, depending on the Confidence Factor. Here we using Mining weka tool and visualization to observe the different item sets and identify the most frequent search items.

II. AN OVERVIEW OF APRIORI and WEKA TOOL

Apriori is designed to operate on database containing transactions (for example, collections of items bought by customers, or details of a website frequented). Each transaction is seen as a set of items (an itemset). Given a threshold $C'$, the Apriori algorithm identifies the item sets which are subsets of at least $C'$ transactions in the database.

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time, and groups of candidates are tested against the data. Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length $k$ from item sets of length $k - 1$. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent $k$-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

III. WEKA TOOL

WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. WEKA is a state-of-the-art facility for developing machine learning (ML) techniques and their application to real-world data mining problems. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms for data preprocessing, classification, regression, clustering, association rules; it also includes a visualization tools. The key features of weka are it provides many different algorithms for data mining and machine learning and it is is open source and freely available and platform-independent and it provides flexible facilities for scripting experiments.

The processing steps in weka includes first we collect the sample data min Excel file formats and after that it is converted into .arff format and after it is again converted in to .csv format that one is used in weka explorer to analyze apriori and visualization methods.
III. METHODOLOGY

The main methodology used for this paper was Data Mining Aprior Algorithm and weka tool for study and analyze the frequent item search patterns through the survey of journals and publications in the field of Data mining, computer science and engineering and the research focused on more recent publications

IV. EXPERIMENTAL RESULTS

4.1 Aprior Algorithm-Frequent Item Set Pruning

In this paper we collect a sample data of 60 students their searched items in shopping portals flipcart, amazon, snapdeal, etc., here we fix the minimum supported confidence factor 5 for Apriori manual and 0.9 for Automation weka tool.

<table>
<thead>
<tr>
<th>Item</th>
<th>Support</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dress</td>
<td>60</td>
<td>0.9</td>
</tr>
<tr>
<td>Books</td>
<td>55</td>
<td>0.8</td>
</tr>
<tr>
<td>Jeans</td>
<td>50</td>
<td>0.7</td>
</tr>
<tr>
<td>Shoes</td>
<td>45</td>
<td>0.6</td>
</tr>
<tr>
<td>Bag</td>
<td>40</td>
<td>0.5</td>
</tr>
</tbody>
</table>

In this paper we collect a sample data of 60 students as shown in the fig(1), this data contains frequent searching items in online shopping portals like flipkart, amazon, snapdeal...by using this data we generate list items and candidate key generations, here we are taking minimum support confidence factor 5.

RESULTANALYSIS

In this experiment we generate successive list items by pruning items depending on confidence factor minimum support count value 5, finally we identify the the most frequent search item {dress, mobile, watches}.

4.2 Visualization and Frequent Patterns Using Weka Tool

In this study paper the same sample data deployed in to weka tool by changing the file format into .csv format and again converts into .arff format, this .arff format data accessed by weka explorer and generate visualization results. We applying associate apriori algorithm with confidence factor 0.9 it identify the frequent search item {dress, watch, jewellery}.

<table>
<thead>
<tr>
<th>Run information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scheme:</td>
</tr>
<tr>
<td>Relation:</td>
</tr>
<tr>
<td>Instances:</td>
</tr>
<tr>
<td>Attributes:</td>
</tr>
</tbody>
</table>
item3

=== Associator model (full training set) ===

Apriori

Minimum support: 0.1 (6 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 9
Generated sets of large itemsets:
Size of set of large itemsets L(1): 4
Large Itemsets L(1):
item1=dress 29
item2=watches 8
item2=dress 10
item3=jewellery 11
Size of set of large itemsets L(2): 2
Large Itemsets L(2):
item1=dress item2=watches 7
item1=dress item3=jewellery 7

V. CONCLUSION

We conclude that in this paper study how to analyze frequent item search patterns by using visualization methods and Apriori Algorithm and with the help of weka tool. This type of analysis will be helpful for more research groups and publishers to study and introduce new innovative methods.

VI. REFERENCES


