Abstract:
E-commerce business constantly decides innovative strategies to increase their sales and hence earn profit. They mainly strive to boost the sale of those items that are rarely purchased. There are few borderline-rare items that lie just below them in minimum support threshold and may have a strong correlation with frequent items. The minimum support threshold is the user defined minimum support value for an item. If these borderline rare items are strategically placed in the market then it can help the e-commerce industry to improve their sales further. In this paper, we propose a hybrid approach, MSD-Apriori to discover borderline-rare elements which are below but close to minimum support threshold and have strong correlation with frequent items. The hybrid approach is formed by integrating MS Apriori with Dynamic Apriori. MS Apriori finds the borderline-rare item sets from the web logs and Dynamic Apriori discovers those items among these that share strong correlation with the frequent items by association rule mining. The proposed method is evaluated on Kosarak, a real data-set that gives encouraging results.

I. INTRODUCTION

In online shopping as well as in real time shopping there arises an issue where the business entity needs to know the items that are being sold frequently. Not only the items frequently sold, but other items bought with that frequently bought item (Customer buying behavior). This information is required in order to deal with the imports of items by the business entity depending on the sales record as well as placements of items in the Shopping Malls which increase sales. Several core techniques that are used in data mining describe the type of mining and data recovery operation. The techniques are,11

1. Association Rule: Association (orrelation) is probably the better known and most familiar and straightforward data mining technique. Here, we make a simple correlation between two or more items, often of the same type to identify patterns.
2. Clustering: Clustering is done using one or more attributes by identify in ga cluster of correlating results. Clustering is useful to identify different information because it correlates with other examples so we can see where the similarities and ranges agree.
3. Sequential Patterns and Prediction: Sequential patterns are a useful method for identifying trends, or regular occurrences of similar events. For example, with customer data we can identify that customers buy a particular collection of products together at different times of the year. In a shopping basket application, we can use this information to automatically suggest that certain items be added to a basket based on their frequency and past purchasing history.
4. Prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation.

II. LITERATURE SURVEY

Retail marketers are constantly looking for ways to improve the effectiveness of their campaigns. One way to do this is to target customers with the particular offers most likely to attract them back to the store and to spend more time and money on their next visit. In existing e-commerce applications such as “flipkart.com”, “myntra.com”, “amazon. in”, “ebay.com”, “snapdeal.com” etc. we have many services which helps customer shopping.

The existing services are listed below:
• Filtering Option: In this user can filter the products based on brands, color, price, size etc. This is one of the feature that helps the customer shopping.
• Displaying Rating: Here users can view the ratings and reviews posted by the members of the application which shows the product quality, price level etc.
• Displaying Similar Products based on user visits.
• Display the customer buying patterns.

Limitations of the Existing System:
• Less user satisfaction
• Time consuming in searching products
• Displaying all patterns for all kind of customers
• Less Reliable
• Less Efficient

III. APPROACH

Proposed system is an e-commerce application accessed worldwide. Proposed system increases sales and hence increases profits. Proposed system aim is to boost the sale of those items that are rarely purchased. Proposed system recommends rare items that may have a strong correlation with the frequent items.
System uses hybrid approach (association rules) which includes MS Apriori and Dynamic Apriori algorithms. Border line rare items discovered have the potential to raise sales of an e commerce website. System is an e commerce website where customers can browse products, add products into cart and can place orders and can make payments. System places borderline rare items in the market and then it can help e commerce industry to improve their sales. The features are

1. To build a real time E-Commerce application.
2. Proposed system aims at implementing the recommendation system for consumers to find the product they want.
3. The main strive is to boost the sale of those items that are rarely purchased.
4. To find the correlation between frequently purchased products with rarely purchased products.
5. Proposed system makes use of data mining techniques for the product recomendation.
6. Proposed recommendation system mainly consists of 3models namely–User Model, Recommended Model and Recommenda
tion Algorithm.
7. Proposed system satisfies the consumers to a better extent.

Approach Adapted: According to Software Engineering the approach adopted to develop this project is the Iterative waterfall Model. The iterative waterfall Model is a systematic approach that begins at the feasibility study phase and progress through analysis, design, coding, testing, integration and maintenance. Feedback paths are there in each phase to its preceding phase as show in the fig to allow the correction of the errors committed during a phase that are detected in later phase.

IV. SYSTEM DESIGN

The purpose of the design phase is to plan a solution of the problem specified by the requirements document. This phase is the first step in moving from the problem domain to the solution domain. In other words, starting with what is needed; design takes us to ward how to satisfy the needs. The design of a system is perhaps the most critical factor affecting the quality of the software; it has a major impact on the later phases particularly testing and maintenance.

A. Input design
Input design is the process of converting user-oriented input to a computer based format. Input design is a part of overall system design, which requires very careful attention. Often the collection of input data is the most expensive part of the system. The main objectives of the input design are
• Produce cost effective method of input
• Achieve highest possible level of accuracy
• Ensure that the input is acceptable to and understood by the staff.

B. Data base design
Introduction to SQL Server 2005: Microsoft SQL Server is a full-featured relational database management system (RDBMS) that offers a variety of administrative tools to ease the burdens of database development, maintenance and administration.

C. Output design
Outputs from computer systems are required primarily to communicate the results of processing to users. They are also used to provide a permanent copy of these result for latter consultation. Computer output is the most important and direct source of information to the users. Designing computer output should proceed in an organized well throughout the manner. The right output must be available for the people who find the system ease of use.

D. Architecture design
Architecture focuses on looking at a system as a combination of many different components, and how they interact with each other to produce the desired result. The focus is on identifying components or subsystems and how they connect. In other words, the focus is on what major components are needed.

Three tier Architecture: This consists of three layers.
• The Data Layer: The key component to most applications is the data. The data has to be served to the presentation layer somehow. The data layer is a separate component (often setup as a separate single or group of projects in a .NET solution), whose sole purpose is to serve up the data from the database and return it to the caller.

Through this approach, data can be logically reused, meaning that a portion of an application reusing the same query can make a call to one data layer method, instead of embedding the query multiple times. This is generally more maintainable.
• Business Layer: Though a website could talk to the data access layer directly, it usually goes through another layer called the business layer. The business layer is vital in that it validates the input conditions before calling a method from the data layer. This ensures the data input is correct before proceeding, and can often ensure that the outputs are correct as well.

This validation of input is called business rules, meaning the rules that the business layer uses to make “judgments” about the data. One of the best reasons for reusing logic are that applications that start off small usually grow in functionality. The business layer helps move logic to a central layer for “maximum reusability.”
• Presentation Layer: The ASP.NET web site or windows forms application (the UI for the project) is called the presentation layer. The presentation layer is the most important layer simply because it’s the one that everyone sees and uses.

Even with a well-structured business and data layer, if the presentation layer is designed poorly, this gives the users a poor view of the system. Here is an architecture diagram showing the flow of activities in the system.

E. Algorithm
An algorithm is an effective method that can be expressed within a finite amount of space and time and in a well-defined...
formal language for calculating a function. Starting from an initial state and initial input (perhaps empty), the instructions describe a computation that, when executed, proceeds through a finite number of well-defined successive states, eventually producing "output and terminating at a final ending state. The transition from one state to the next is not necessarily deterministic; some algorithms, known as randomized algorithms, incorporate random input.

MSD Apriori Algorithm
STEP 1: Scan the data set and determine the support(s) of each item.
STEP 2: Generate L1 (Frequent one item set).
STEP 3: Use Lk-1, join Lk-1 to generate the set of candidate k-item set.
STEP 4: Scan the candidate k item set and generate the support of each candidate k-item set.
STEP 5: Add to frequent item set, until C=Null Set.
STEP 6: For each item in the frequent item set generate all non empty subsets.
STEP 7: For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence. Then add to Strong Association Rule.

V. EXPLAINATION
• The item sets taken here are A, B, C, D and E. Assuming that these are the products that are available we will determine the minimum support and minimum confidence for the item sets. Minimum support is equal to the number of records of a particular item set divided by the total number of transactions taken in percentage.
• Similarly minimum confidence is calculated by comparing the relationship between two items divided by the total number of transactions first item set appears.
• For pattern prediction we can consider items that are only equal to or greater than the minimum support and minimum confidence. Hence it is very important to pick the right percentage of minimum support. Else we will end up with combination explosion.
• Here C1 is the candidate 1 item set which will initially be found by finding the support of all the items and compare them with the minimum support that was set previously.
  • Only items that are equal to or greater than the support will be added to L1 list also known as the frequent 1 item set. The frequent one item set (L1) consists of the items that are equal to or more than the minimum support that we found in C1.
  • Based on L1 we find C2 which is the combination of items and similarly compare them to the minimum support and the items eligible that are equal to or more than the minimum support will be added to a new list L2 also known as frequent 2 item set.
  • Now with the combination of the itemset in L2 we obtain a new candidate 3 itemset (C3). We find the support of each itemset and compare it with the minimum support as done previously and add the eligible items to a new frequent 3 itemset L3.
  • Since MSD works on iteration steps it repeats till C is null. As shown in the figure iteration stops at Now we find L which is a combination of all the frequent itemset obtained that is L1, L2 and L3.
  • As shown in the figure each item in L generates an on empty subset. Next we go ahead and find the confidence of the subsets to achieve our final output. This is how we obtain rare itemset from association rule.

VI. TESTING

Testing accomplish the severity of things, but most importantly it measures the quality of the software we are developing. This view pre supposes there are defects in the software waiting to be discovered and this view is rarely disproved or even disputed.

The following types of testing can be automated:
  • Functional - testing that operations perform as expected.
  • Regression-testing that the behavior of the system has not changed.
  • Exception or Negative-forcing error or conditions in the system.
  • Stress - determining the absolute capacities of the application and operational infrastructure.
  • Performance - providing assurance that the performance of the system will be adequate for both batch runs and online transactions in relation to business projections and requirements.
  • Load-determining the points at which the capacity and performance of the system become degraded to the situation that hard ware or software upgrades would be required.

Promote these borderline-rare items with the correct frequent itemset to its users, and hence increase their sale probability. Proposed system uses MSD apriori algorithm to find the relationship between rarely purchased and frequently purchased products. For future enhancement we have many algorithms in data science to find patterns such as eclat algorithm, SFIT algorithm, AIT algorithm etc.

In the future we can add more algorithms to find patterns and algorithms can be compared to find the efficient algorithm. In the project only few categories of products considered for recommendations such as men’s clothing, women’s clothing, kids clothing, mobiles, laptops, tabs, Men’s Appetures and footwear, in future we can add more product categories for recommendations. We can also add visitor query module, where visitors can post queries to administrator and admin can send reply to those queries.

<table>
<thead>
<tr>
<th>TC#</th>
<th>Description</th>
<th>Expected Result</th>
<th>Actual Result</th>
<th>Status of Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>TC01</td>
<td>Execute run the application</td>
<td>Application should run without any interrupts.</td>
<td>Application is executing properly</td>
<td>Pass</td>
</tr>
<tr>
<td>TC02</td>
<td>Verification of Admin Login Input User Name and Password then click on Login button.</td>
<td>Admin User Name &amp; Password should be check/verify with database.</td>
<td>Admin User Name &amp; Password successfully checked with database.</td>
<td>Pass</td>
</tr>
<tr>
<td>TC03</td>
<td>Verification of Input User Name &amp; Password of Admin.</td>
<td>Admin User Name &amp; Password is valid then it should navigate to respective Admin home page.</td>
<td>Admin User Name &amp; Password is valid then successfully navigating respective home page.</td>
<td>Pass</td>
</tr>
<tr>
<td>TC04</td>
<td>Verification of Input User Name &amp; Password of Admin. (Invalid Case)</td>
<td>Admin User Name &amp; Password is invalid then show message that Input Username &amp; Password is wrong.</td>
<td>If User Name &amp; Password is not valid or wrong input then message box shown that User Name &amp; Password wrong.</td>
<td>Pass</td>
</tr>
</tbody>
</table>

VII. CONCLUSION

Every business objective is to earn best profits. Current e-commerce applications such as “flipkart.com”, “amazon.com”, “myntra.com” etc. recommends only frequently purchased products but doesn’t concentrates on rarely purchased products. Attracting customers and promoting rarely purchased products is a factor to improvise over all business profits. Finding the borderline-rare items and the sequences of frequent items with which they are correlated so that the E-Commerce websites can

VIII. REFERENCES

[2]. YUN SING KOH, SRI DEVIRAVANA, “Unsupervised Rare Pattern Mining: A Survey” in ACM Transactions on Knowledge Discovery from Data, Vol. 10, No. 4, Article 45,2016

[6]. Bing Liu et al., “Mining Association Rules with Multiple Minimum Support in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-99), August 15-18, 1999, San Diego, CA, USA