A Novel Approach of Product Recommendation Using Utility-Based Association Rules

Stuti Stuti, Jaypee Institute of Information Technology, India

Kanika Gupta, Jaypee Institute of Information Technology, India

Nishant Srivastava, Jaypee Institute of Information Technology, India

Ankita Verma, Jaypee Institute of Information Technology, India

ABSTRACT

An exorbitant source of data is easily available, but the actual task lies in using this data efficiently. In this article, the aim is to analyse the significant information embedded in the customer purchase behaviour to recommend new products to them. The proposed scheme is a two-fold approach. First, the authors retrieve various product correlations from the vast library of user transactions. Based on these product correlations, utility-based association rules are learned which depict the customer purchase behaviour. These rules are then applied in a recommender system for novel product suggestions to the customers. With improved utility-based mining, the paper tries to incorporate the usefulness of an item set like cost, profit, or any other factor along with their frequency. In this paper, the authors have deployed the rules discovered from both the conventional frequent item set mining and improved utility-based mining on an e-commerce platform to compare the accuracy of the algorithms. The obtained results establish the efficacy of the proposed algorithm.

KEYWORDS

Customer Relationship Management, Market Basket Analysis, Recommender System, Utility Mining

INTRODUCTION

Analyzing purchase behavior among various consumer choices with respect to numerous products and services during a single shopping experience to realize possible correlations has been a popular method in order to take the shopping experience on a whole new level. This technique is also popular as association rule-mining due to its involvement in detecting hidden patterns from huge transactional database through the extraction of associations or co-occurrences in transactional data (Kim et al., 2012). In a study by Mostafa (2015), novel data mining technique is applied for the first time to investigate consumer behavior in Kuwait. Data mining has always played a crucial role in Customer Relationship Management helping with the retrieval of sensible information from large sets of data only to further improve the management of relationship.

A traditional problem, as argued by Burke (1996), among retailers is the determination of products' group by brand or product type. Through the method of identifying consumer purchase behavior by obtaining linkages or co-occurrences from stores' transaction logs Market Basket Analysis helps in

strategically designing a layout such that products associated with each other are juxtaposed. Several other marketing applications like cross-selling (e.g., Russell and Petersen, 2000), designing promotion campaigns (e.g., Abraham and Lodish, 1993) are resolved with the help of Market Basket Analysis. A-priori algorithm (Agrawal et al., 1993), one of the most prevalent methods till date to obtain associations from transactions uses the concept of minimum support, assessing the frequency of both P and Q in the transactions, and minimum confidence, assessing the correctness of the rule, only to imply that a pattern of P Q means if P is bought then Q will be bought as well. However, A-priori (Agrawal et al., 1993) fails to consider the utility because of its prime concern on frequency. The term utility can be defined as usefulness of the presence of item sets in transactional logs, and is expressed as a quantity in terms of profit or sales, etc. of an item (Srivastava et al., 2018). The limitations of existing frequent item mining techniques have been overcome through the amalgamation of both high frequency and high utility while determining useful associations (Srivastava et al., 2018).

The associations discovered from a plethora of data has many applications. Recommendation Engine being another one of them. People are unaware of their own requirements until they are told so. A study found out that purchase behavior of shoppers who experience personalization, ranging from an influenced future purchase to an immediate added order, is amenable. A system to filter information aiming to analyze and estimate the 'rating' that user would give to an item and accordingly make recommendations is one of the ways to personalize the shopping experience. In a study by Hallowell (1996), there was a correlation between customer satisfaction and customer retention, and between customer retention and profitability. The overall effects, including the before and after effects, of a recommendation system combined with pure search (querying) and browsing (directed or non-directed), they allow users facing a huge amount of information to navigate that information in an efficient and satisfying way (Davidson et al., 2010).

The aim of a recommender system is to produce a list of possible inclinations for a user on the basis of his/her previous interests in different products and services. In general, recommendation engines are categorized into four parts: demographic filtering, collaborative filtering (Manolopoulos et al., 2008), content-based filtering and hybrid fusions (Lopez et al., 2003). Personalized recommendations are a way to go not only in e-commerce websites but also in social networks by helping them make decisions based on their social relations, what blogs to read (Arguello et al., 2008), what software items to use (Carmel et al., 2009) etc., and how to quantify a trade-of between accuracy and privacy (Korolova et al., 2011). Some researchers attribute to rule-based recommendations for a highly personalized experience. Combining the rules generated with a personalized shopping experience is termed as rule-based recommendation system. In this paper, we intend to compare the results of rules discovered from Improved Utility Based Mining (Srivastava et al., 2018) and the traditional method of Frequent Item Set Mining (Agrawal et al., 1993) by implementing them in a rule-based recommender system. When a rule is in the following form: IF antecedent THEN consequent, then a recommendation of consequents is provided according to a consumer's purchase history and behavior i.e. antecedentes.

Summarizing, the key contributions of this article can be enlisted as:

I. Detailed analysis and more experimental study is conducted to establish the efficacy of our prior proposed method Improved Utility Based Mining (Reference).
II. We have developed a recommender system to suggest novel products to the users via the improved utility based association rules.
III. We have established the efficacy of the proposed scheme by empirical analysis conducted on the real-world datasets.

Rest of the article is organized as follows: Section 2 gives the state-of-the-art techniques and some background information. In Section 3, we have given the problem definition formally. Section

4 discusses our proposed scheme in detail. Experimental setup and obtained results are discussed in Section 5. Finally, concluding the article in Section 6.

BACKGROUND

Here, we review some research works relevant to association rule mining, utility mining, rule-based recommendation engines that prove useful in studying the customer purchase pattern.

Association rule mining, one of the most paramount and well researched techniques of data mining, was first discussed by Agrawal et al (1993). The method identified frequent individual items and extending them to supersets till those sets satisfy the minimum required threshold over transactional databases. Frequent Item Set Mining with its coming laid down a path for various other algorithms. One such is FP-tree and an FP- growth method, where rules are extracted without generating candidate sets by introducing a data structure (Han et al., 2000). With additional research in the field and striving for better and more efficient solutions many other algorithms have also followed ever since. An extended prefix tree structure, developing an efficient FP-tree-based mining method to mine the comprehensive set of frequent patterns by pattern fragment growth was proposed by Han et al (2000). Chen et al (2005) proposed an algorithm similar to A-priori for a multi-store condition additionally consisting of data on store (area) and time where the guidelines hold. A high utility mining algorithm was proposed by Bhattacharya et al (2012) depending on the profit associated with item sets satisfying the minimum threshold.

Recommender systems is one of the most fascinating fields in both research and industry finding relevance in numerous discipline, including machine learning, text mining, artificial intelligence, data science, visualization, human computer interaction, and many more. One of the versatile frameworks show an information structure for putting away the found incessant item sets which is particularly appropriate for recommender frameworks. The recommendation calculation uses this information structure to create recommendations effectively progressively, without the need to produce all association rules from visit item sets (Dai et al., 2001). Hu et al (2012) in their paper, based on the research on some existing models and algorithms, aimed to predict the rating for a product that a customer has never reviewed, Item Similarity, Bipartite Projection and Spanning Tree were implemented, based on the data of all other users and their ratings in the system on some existing datasets indicating Spinning Tree has the best result, for old users, and Item Similarity best with mean squared error (MSE), for new users. Bipartite Projection has the best result and in terms of computational performance, Bipartite Projection is the swiftest algorithm.

With a number of limitations tied to various methods it is important that meaningful recommendations based on purchase patterns are generated to be more insightful of customer buying habits as opposed to present methodologies which utilize users' ratings or history. Cakir et al (2012) using association rule mining and the personalization approach through collaborative filtering customized a business site. In a study by Badur et al (2012), using an online dating site and experiments and association rule-based recommendation engine, showed that scores of matching couples predicted by this engine were significantly high. Gatzioura, et al (2015) implemented a methodology using a hierarchical categorization of the items that the transactions consist of, in order to find out the structure of users' preferences and generate valuable recommendations. Even after a decade of research, customer purchase pattern continues to unleash various unknown possibilities. Table 1 recapitulates some related research works.

MOTIVATION AND PROBLEM STATEMENT

Though the field of frequent item set mining has been a heavily researched one ever since its inception, traditional association rule mining is still the most cited and all-prevailing in the field of market basket analysis. However, owing to limitations like viewing all items as equally important, not considering

Table 1. Summary of the related work

Work	Research question	Methodology	Contribution	Limitation	Scope
Mostafa, 2015	An analysis of purchase patterns in Kuwait- Market Basket Analysis.	Useful rules are mined by association rules mining (ARM) and products' positive and negative correlations using the rules.	In coming up with price promotion and marketing strategies.	Purchases done on the same day but not together at the same time are not talked about.	Provides a rich picture of Kuwaiti market helping businesses.
Chen et al, 2005	An extension of A-priori to encapsulate a multiple store environment.	Extended A-priori with each transaction having an additional timestamp and a store identifier.	Useful in dealing with product inventory, and delivery strategies.	Fails to discuss the association rule mining in environments like iterative, online and so on.	Possible expansion by taking constraints like spatial, temporal in mind.
Dai H. et al, 2001	Effective and scalable ways for Web personalization to tackle the lack of explicit user ratings.	Matches the current user's activity against the discovered patterns.	Scalable framework presented.	ARM recommendations- primitive nature.	Can be extended to real time market situations.
Srikant, Vu, & Agrawal, 1997	Generating association rules with item constraints.	Generating rules using a Boolean expression.	Extracting rules according to user constraints.	Multiple taxonomy can lead to conflicting rules.	Facilitates knowledge accumulation and creation.
Murat Efe Aras, Ozgur Cakir, 2012	Offer products, users will find interesting, and get higher converts via recommendation engine by using ARM.	C# programming language and works on .NET framework.	Recommendation system when used in week 3 increases the basket ration when compared to week 1 when it wasn't used.	Random recommendations don't make difference.	Can be extended to real time market situations.

the varying unit costs among items engendered to various other algorithms. Therefore, some research has been done in the field of utility item set mining. But due to the inability of applying Downward Closure Property of A-priori to obtain high utility item sets simply increases the processing time. Hence in the study by Srivastava et al (2018) we tried implementing an algorithm which is an amalgamation of both a-priori and high utility item set mining taking both utility and frequency as factors along with the Downward Closure Property to compute associations.

Market basket domain can be attributed to consisting large number of co-occurring items and transactions. In such cases, predicting whether an item will be liked by a user or not is not the only task but also to encapsulate the presence or absence of an item during the spree, and thereon recommend complementary items accordingly. Most of the current recommendation methodologies do not take into account these co-occurrences while the association rules methodology that has been used as a basis for recommendations in such cases does so (Agrawal et al., 1993). Collaborative Filtering's inability to recommend an item not been previously rated, requirement of sufficient users to make recommendations are some of the limitations (Manolopoulos et al., 2008). Similarly, in Content-Based filtering longer words have higher chance of overlap with the profile, not every word has similar importance are some of the named drawbacks.

An efficient methodology which tries to aggregate all possible limitations discussed above in order to expand the possibilities in the business world is something we tried to achieve in this paper and attempted to study and compare the results of the conventional method of Frequent Item Set Mining proposed by Agrawal et al (1993) and the ensemble algorithm discussed by Srivastava et al (2018) we, in this paper, employ the correlations and rules derived from the two algorithms in the simulated environment of an e-commerce website and propose recommendations through association rule-based recommendation engine.

PROPOSED WORK

Taking the base of the conventional and utility centric methods namely A-priori algorithm and HUIM, we in the study by Srivastava et al (2018) discussed an ensemble algorithm which considers both the frequency of the items present and its utility in terms of the profit associated with it or the revenue it generated on transactional data. We will first present the classical association rules generation algorithm (Agrawal et al., 1993) and then present our previously proposed utility based association rules generation (Srivastava et al., 2018). Our main objective is to systematically study the recommendations generated by both of these algorithms and impact of adding various other utility based information in the association rules mining. In order to compare the performance of A-priori (Agrawal et al., 1993) and Improved Utility Based Mining (Srivastava et al., 2018) we implemented the following algorithms over a data set with 52 lakh entries to generate useful association rules.

For a given transactional database consisting of customer purchases such that each transaction consists of the items purchased by the customer, the algorithm proposed by Agrawal et al (1993) generates different item sets where K varies from 1 to combinations of unique and varying items till no such transaction exists. For every K-Itemset we check the item set with the minimum support and accordingly prune those instances which fail to fulfill the threshold. For every K-Itemset, support specifies how frequently the items in the set appear in the database of the transactions. Finally, by the end of the scan only the strong rules which satisfy the minimum support are left. The pseudo-code of the algorithm is given as:

```
Frequent Item Set Mining Algorithm
Input: Transactional database
Output: Association Rules
TID: Transaction ID
S: minimum support
Algorithm:
1.      ∀TID
            set K=1
generate K-Itemset candidates
2.    ∀X ∈ K-Itemset
1.      check minimum support
2.      if Sup(X)< S
                    prune(X)
            else
                    X ∈ Frequent Itemset
3.   Set K=K+1
            if K-Itemset=Ø
                    Stop
            else
                    Repeat steps 1 and 2
```

The improved utility-based algorithm first aims at generating transactional utility using the product of frequency and utility (cost in this case) for each transaction. Then, the transaction weighted

utilization is calculated using different combinations of items and pruning the sets not satisfying the utility threshold. Once the various scans of pruning are completed strong association rules are generated as a result. The complete algorithm is given as follows:

```
Improved Utility Based Mining
Input: Transactional database, Utility (cost) of each item
Output: Association Rules
TID: Transaction ID
F: frequency of each item in a transaction
C: cost of each item
E: user specified threshold
Ē: utility threshold
HU: High Utility Item Set
HTWU: High Transaction Weighted Utilization Item Set
Algorithm:
1.       ∀TID
calculate transactional utility using  F X C
2.       ∀X ∈ itemset
calculate transaction weighted utilisation (twu)
3.       If twu(X)>E
X ⊆ HTWU
4.       If E = Ē → HU ⊆ HTWU
```

Once the implemented algorithms generate useful associations, we tried to analyse the market, its trends and associations between products through the various purchases and expenditures done by the customers in order to give them the most personalized experience possible during their next turn through recommendations using association rules gathered through the following algorithm where in for every search query provided if the product is present in the product data then a list of result is provided. In the result if an item is selected then it is matched for antecedents accordingly the consequent for the same is recommended. Figure 1 shows the flow diagram of the implemented work.

```
Recommendation Engine Algorithm
Input: Search Query (Q)
Output: Recommendation using Association Rules
X: Product
Q: Search Query
AR: Association Rules
Algorithm:
1.       If Q=true
Result=∀P ∈ Product Data
2.       ∀Result
   If X ∈ Result is selected
1.       match (X, A→B) such that R1=A,∀X∈ Association Rules
such that X is the antecedente
2.       retrieve AR(X)
3.       ∀AR(X)
   If support(AR(X)) =max OR utility(AR(X))=max
3.       recommend the consequents of A→B where A=X i.e. B
```

In order to provide a platform to implement the rules a recommendation engine was designed. A directory of the rules retrieved was then fed into the database along with various other tables like product information, customer order, and cart items in order to simulate an e-commerce like environment. Some added functionalities (refer Figure 2) of the e-commerce website are viewing product details, rating, commenting, and so on.

The mechanism for the simulated e-commerce website is as follows (refer Figure 3):

1. Enter login credentials
2. Browse or search items
3. View product details along with recommendations
4. Add items to cart as per the choice
5. Update, View or Checkout cart

The customers were recommended the corresponding consequents on the basis of their searches and items in their cart which we in this paper assume to be the antecedents i.e.

```
If    P→Q then
      P: antecedent
                Q: consequent
```

These recommendations were further sorted according to their respective support, confidence and utilities.

EMPIRICAL ANALYSIS

In this section, we will first describe the dataset used and then we will present as well as systematically analyze the obtained results.

Dataset Description

A transactional dataset (Chen et al., 2012) containing transactions for a UK-based and registered retail of unique all-occasion gifts. The dataset consists of the following attributes:

- Invoice number
- Stock code
- Description
- Quantity
- Invoice date and time
- Unit Price
- Customer number
- Country of transaction

A total of 541,909 instances are a part of this table where a group of k instances imply a single transaction and leading to about 5000 transactions.

First of all, we will present the analysis of the rules generated by the proposed utility based algorithm of association rule generation and then we will comparatively analyze the recommendations generated by these rules to the recommendation generated by the A-priori algorithm.

Analysis of the Association Rules Generated

First, we will analyze the association rules generated by our proposed utilty based mining algorithm (Srivastava et al., 2018) from the above-mentioned data set.

In order to study the intricacies of the rules fetched from the implementation of Improved Utility Based Mining (Srivastava et al., 2019) the following measures were used for a rule of the form A→B where A and B are disjoint item sets such that $A \cap B = \varnothing$. Here $frequency(A)$ gives the frequency of occurrence of the items in the set A in all the transactions, N gives the total number of transactions in the database.
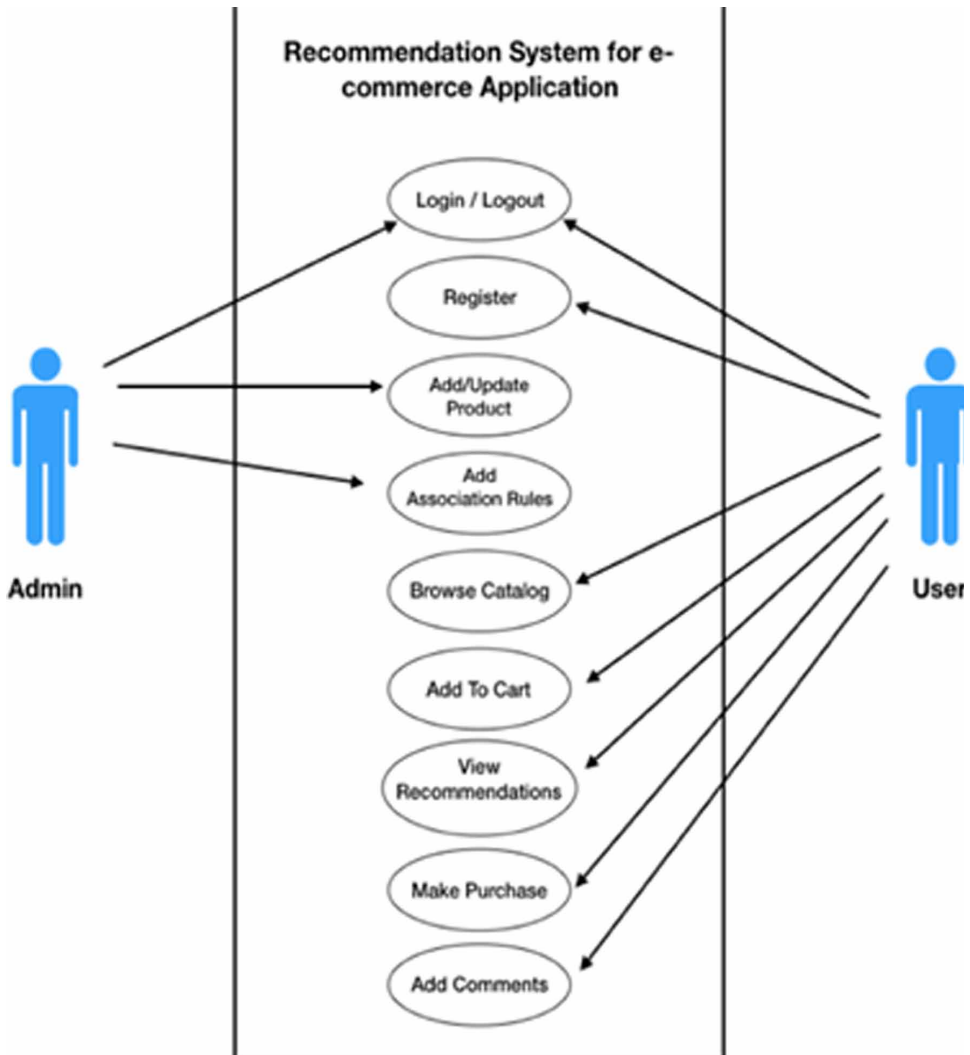
Figure 1. Control flow diagram
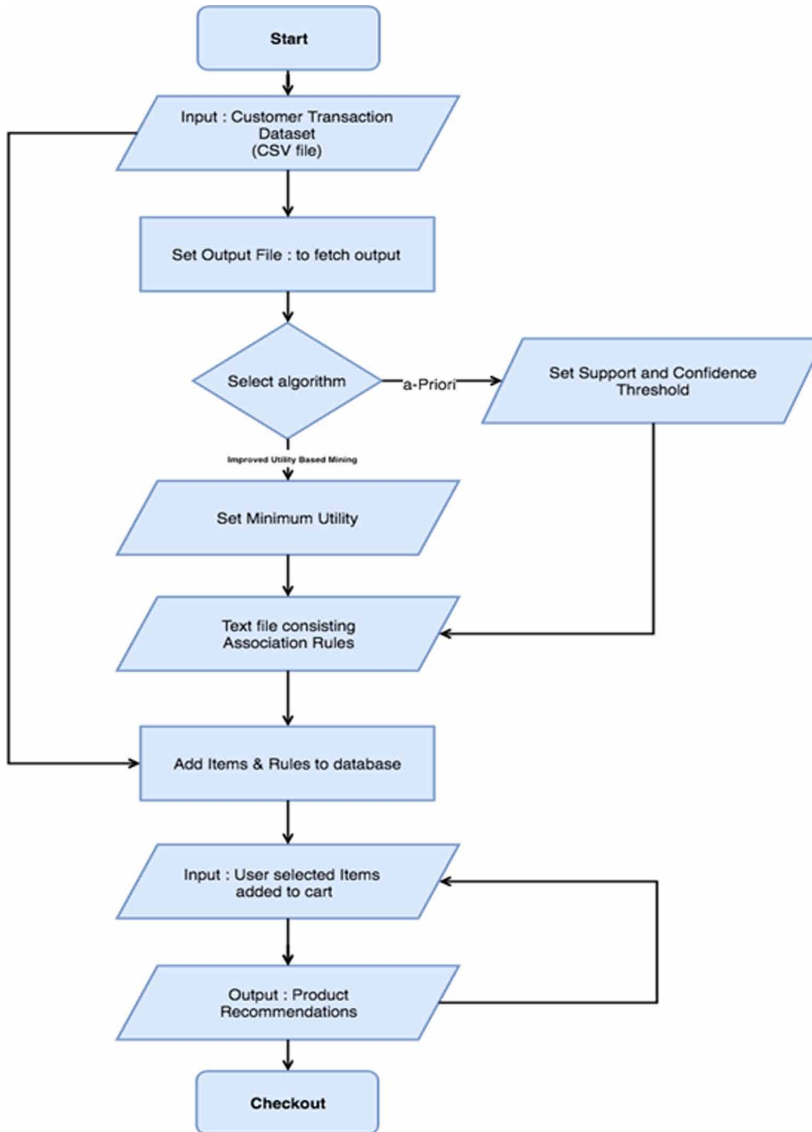
- Support is an indication of how frequently the items in the set $A \cup B$ appear in the database.

$$Support = \frac{frequency\left(A \cup B\right)}{N} \tag{1}$$

- Confidence indicates the number of times the if and then statements have been found to be true in the given set of transactions.
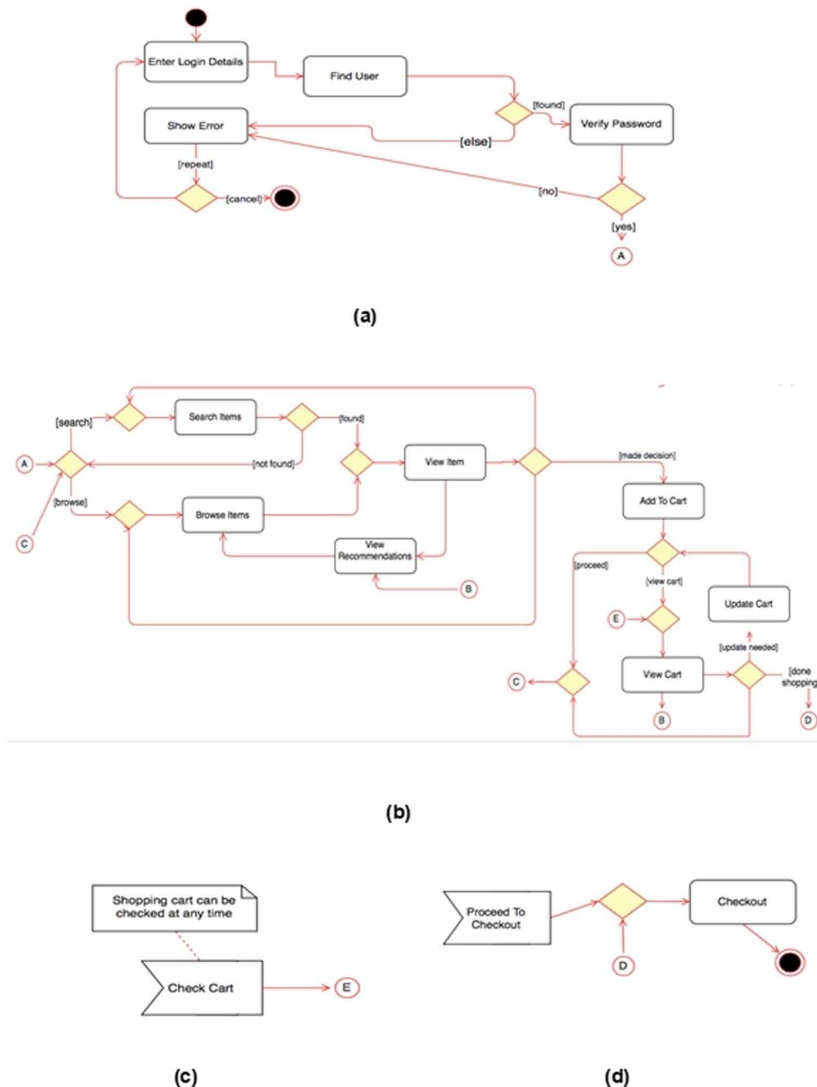
Figure 2. Use case diagram

$$Confidence = \frac{frequency\left(A \cup B\right)}{frequency\left(A\right)} \tag{2}$$

- Lift value is a measure of importance of a rule. A lift greater than 1 signifies that the antecedents and consequents are positively correlated.

$$Lift = \frac{frequency\left(A \cup B\right)}{frequency\left(A\right) * frequency\left(B\right)} \tag{3}$$

Figure 3. Activity diagrams of recommendation engine depicting different activities performed

Support and confidence are often expressed as values ranging from 0 to 1. The lift is a value between 0 and infinity.

Figure 4 gives a peek at the lift and confidence measures of some of the rules. The correlation between the variables support, confidence and lift was found to be varying for the rules mined. In Figure 5(a) we see that rules with high lift have low support. Bayardo, Jr. and Agrawal (1999) in their study found that rules on the support or confidence border are the most interesting ones. In Figure 5(a) similar plot is used to check the rules' noteworthiness. Similarly Figure 5(b) depicts that the relationship between order and support is an inverse one (Seno and Karypis, 2005). Order in this plot is used for shading and studying the varying relationship between support and confidence of the rules generated.

A 3-dimensional matrix is used and the interest measure is represented by 3D bars at the intersection as shown in Figure 6 (Wong, Whitney, and Thomas 1999; Ong et al. 2002). Number of

Figure 4. LIFT and confidence measure of rules

```
> inspect(sort(mbaUM$model,by="confidence")[1:5])
    lhs                    rhs      support    confidence lift      count
[1] {23286}        => {23287} 0.03116883 1.0000000  25.666667 12
[2] {22698,22699}  => {22697} 0.02597403 1.0000000  19.250000 10
[3] {20726,20727}  => {20725} 0.02597403 1.0000000  10.694444 10
[4] {22698}        => {22697} 0.02857143 0.9166667  17.645833 11
[5] {23209,85099B} => {23203} 0.02857143 0.9166667   8.607724 11
> inspect(sort(mbaUM$model,by="lift")[1:5])
    lhs                    rhs      support    confidence lift     count
[1] {23286}        => {23287} 0.03116883 1.0000000  25.66667 12
[2] {23287}        => {23286} 0.03116883 0.8000000  25.66667 12
[3] {21932}        => {21933} 0.02597403 0.8333333  24.67949 10
[4] {21933}        => {21932} 0.02597403 0.7692308  24.67949 10
[5] {22697,22699} => {22698} 0.02597403 0.7142857  22.91667 10
```

rows or columns depends on the variety of unique item sets in the consequent or antecedent in the set of rules.

Graph-based techniques (Buono and Costabile, 2005; Rainsford and Roddick 2000; Klemettinen, Mannila, Ronkainen, Toivonen, and Verkamo 1994; Ertek and Demiriz 2006) help in visualising association rules with vertices and edges giving vertices item labels to represent items, and item sets or rules are regarded as a second set of vertices. Plot in Figure 7 represents items as vertices and rules connecting them with directed edges in two scenarios. 7(a) shows graph visualization for 10 rules and 7(b) shows graph visualization for 100 rules. This representation focuses on how the rules are composed of individual items and shows which rules share items. In Figure 8, we see parallel coordinates plot for 10 rules where an arrow's width depicts support and colour intensity tells the confidence. As the number of rules increase the number of crossovers also between the lines increases Yang (2003).

Analysis of the Recommendations Made Via Association Rules

Efficiency comparison of the recommendations made by A-priori and Improved Utility Based Mining algorithms are done using the following metrics:

- Precision is the ratio of retrieved items that are relevant and all the retrieved items.

$$Precision = \frac{\left(Relevant \cap Retrieved\right)}{\left(Retrieved\right)} \tag{4}$$

- Recall is the ratio of relevant items retrieved and all items that are relevant to us.

$$Recall = \frac{\left(Relevant \cap Retrieved\right)}{Relevant} \tag{5}$$

- F-score is the single measure which comprehensively coalesces the precision and recall score, thus giving us a single measure for the same i.e. the mean of precision and recall.

Figure 5. (a) Scatterplot between support, lift and confidence (b) Scatterplot between support, confidence and order
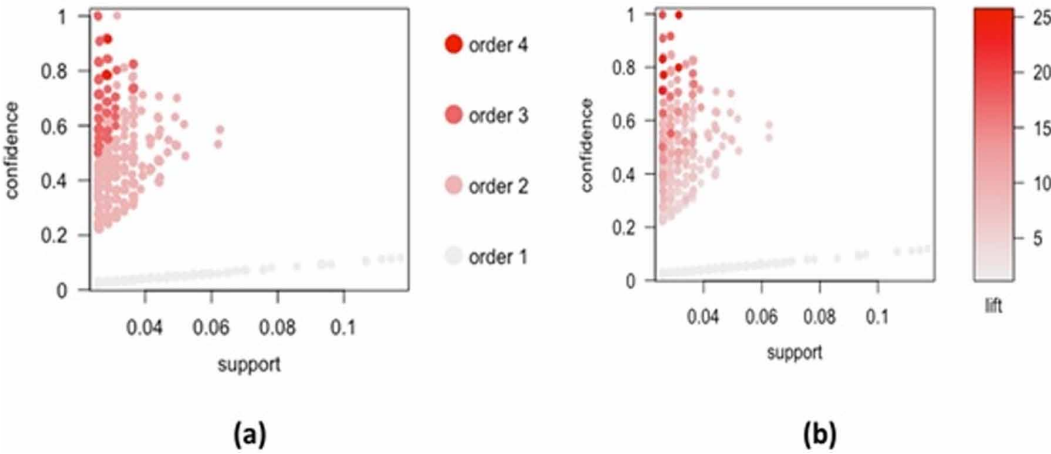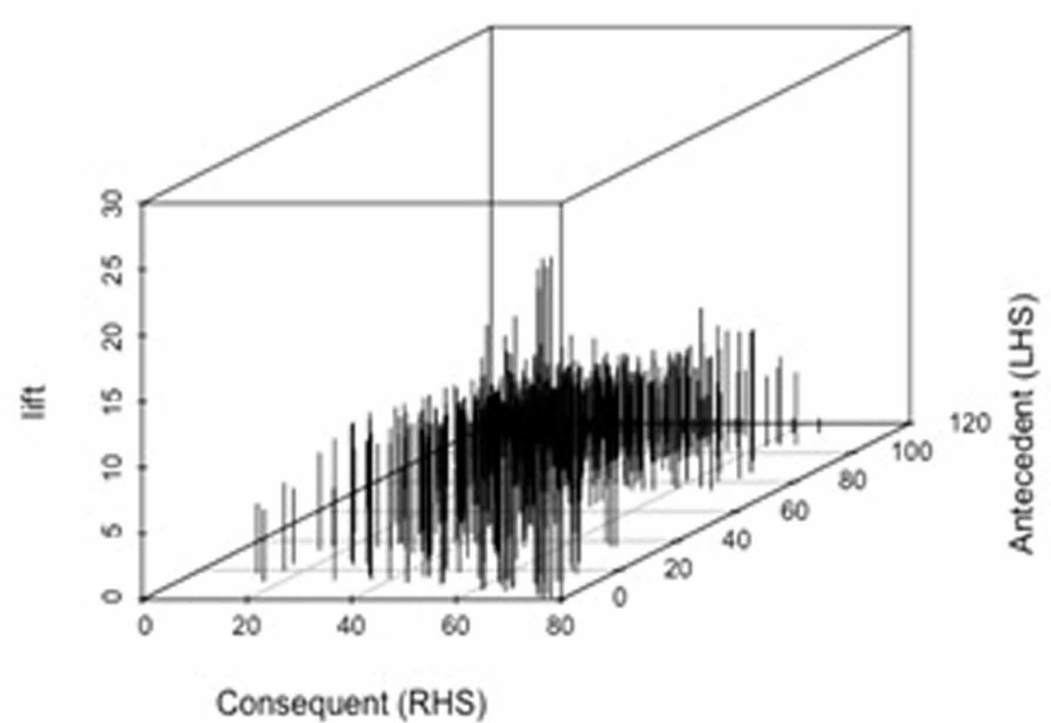

(a)   (b)

Figure 6. Matrix visualization

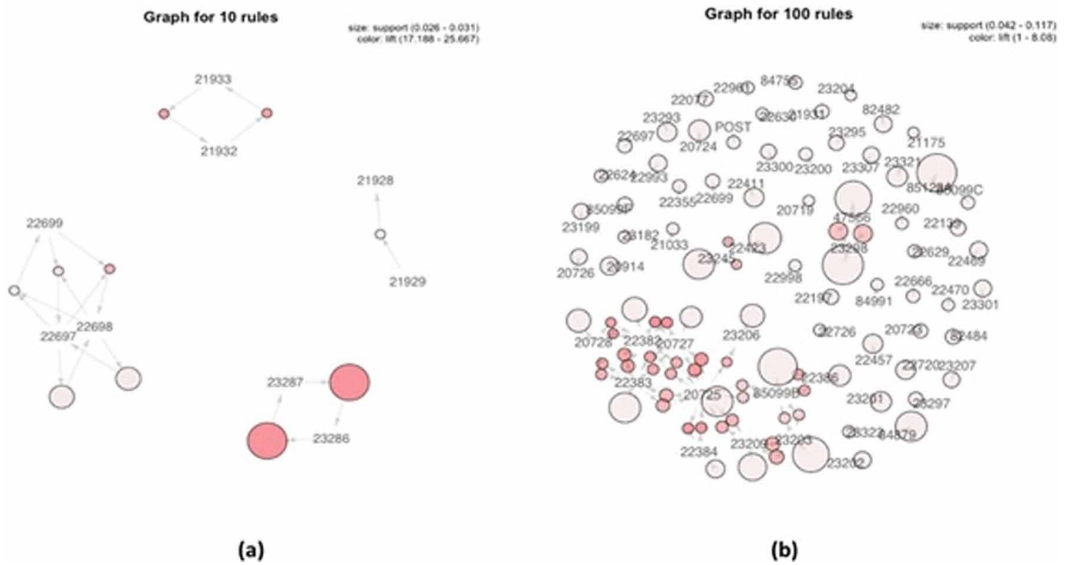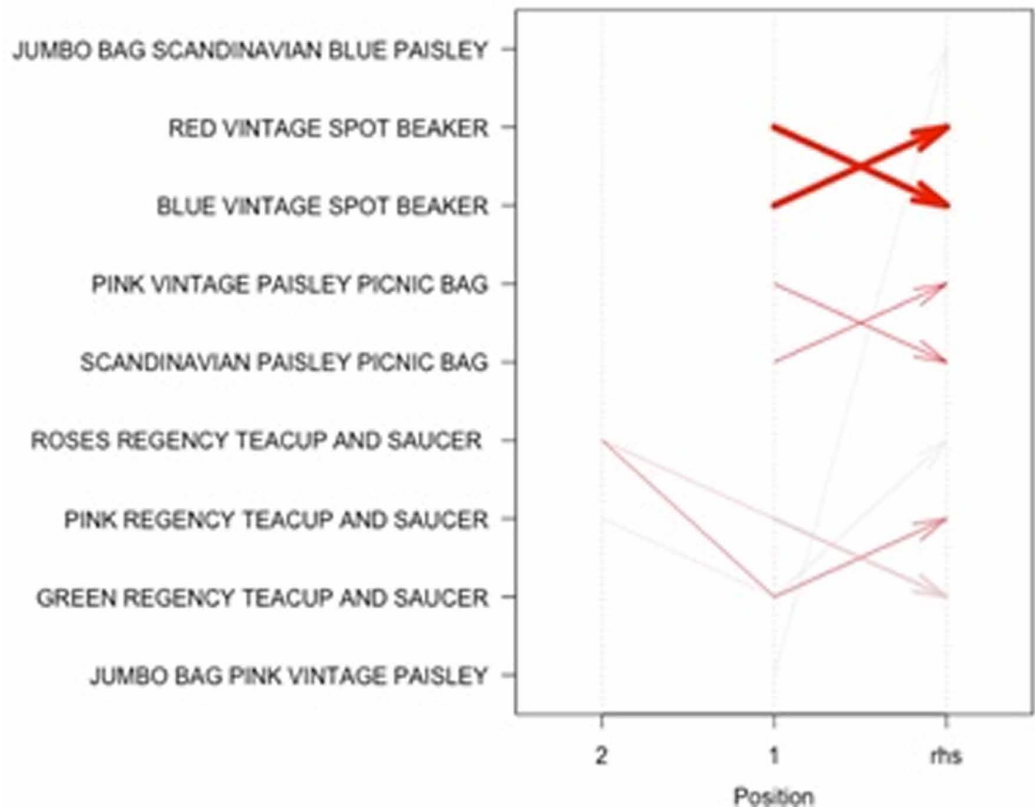**Figure 7. Graph visualizations (a) for 10 rules (b) for 100 rules**

(a)

(b)

Figure 8. Parallel coordinates plot for 10 rules

$$f - score = \frac{2 * \left(Precision * Recall \right)}{\left(Precision + Recall \right)} \hspace{3cm} (6)$$

- Receiver Operating Characteristic (ROC) curve is a performance analysis plot of the recommendation engine between true positive rate (TPR) and false positive rate (FPR). The true positive rate is also called Recall or Sensitivity and FPR is also called as (1-specificity). False Positive Rate (FPR) specifies the rate of irrelevant items which are identified as irrelevant items.

In the generated rules implementation over a recommendation engine of A-priori and Improved Utility Based Mining we divided our data in training and testing set. The training set was used to generate rules and in the testing set within transactions, we deleted items accordingly to check if the recommended items are congruous to the rules retrieved. By employing the above mentioned strategy, we have created different versions of test and train datasets from the complete dataset and each of such test scenorio are denoted with the help of test case id.. Table 2 gives the performance of the recommender systems developed on the basis of both the algorithms. In the table, bold faces denotes the better performance for the particular test case. From the table, it can be concluded that our algorithm outperforms the other one in almost all the test scenarios.

Table 2. Performance of the algorithms for recommendation

Test Case ID	Precision		Recall		Fscore	
	A-priori	*Utility Based algorithm*	*A-priori*	*Utility Based algorithm*	*A-priori*	*Utility Based algorithm*
1	0.45	**0.61**	0.78	**1**	0.57	**0.76**
2	**0.60**	0.52	0.82	**0.95**	**0.69**	0.67
3	0.52	**0.70**	0.74	**0.90**	0.61	**0.77**
4	0.58	**0.61**	0.76	**0.98**	0.66	**0.75**

The trade-off between precision and recall for different test cases is shown in Figure 9 by a Precision-Recall curve. A high area under the PR curve signifies a greater recall and precision i.e, a low false positive rate and low false negative rate. The other graphical analysis of the recommender system's performance is ROC curve as shown in Figure 10. The closer the graph is to the top left border, the more accurate the test. A greater AUC is a measure of accuracy. ROC curve is an effective method helping in evaluation of the quality and performance of the algorithms. An overlapping Precision-Recall curve and ROC curve give us a clear perception of the area covered by both the curves and by both the conventional algorithm of Frequent Item Set Mining and the amalgamated algorithm of Improved Utility Based Mining consisting of both the frequency of the item in the transaction and the corresponding utility in the form of profit, unit price or any other utility measuring factor.

In Table 3, we have included the average values of the three measures upon which we have compared the two algorithms, viz Precision, Recall, F-score and ROC curve over all the designed test scenarios. From the table 3, following conclusions can be made:

Figure 9. Overlapping PR Curves: Orange for A-priori, Red for Utility based algorithm
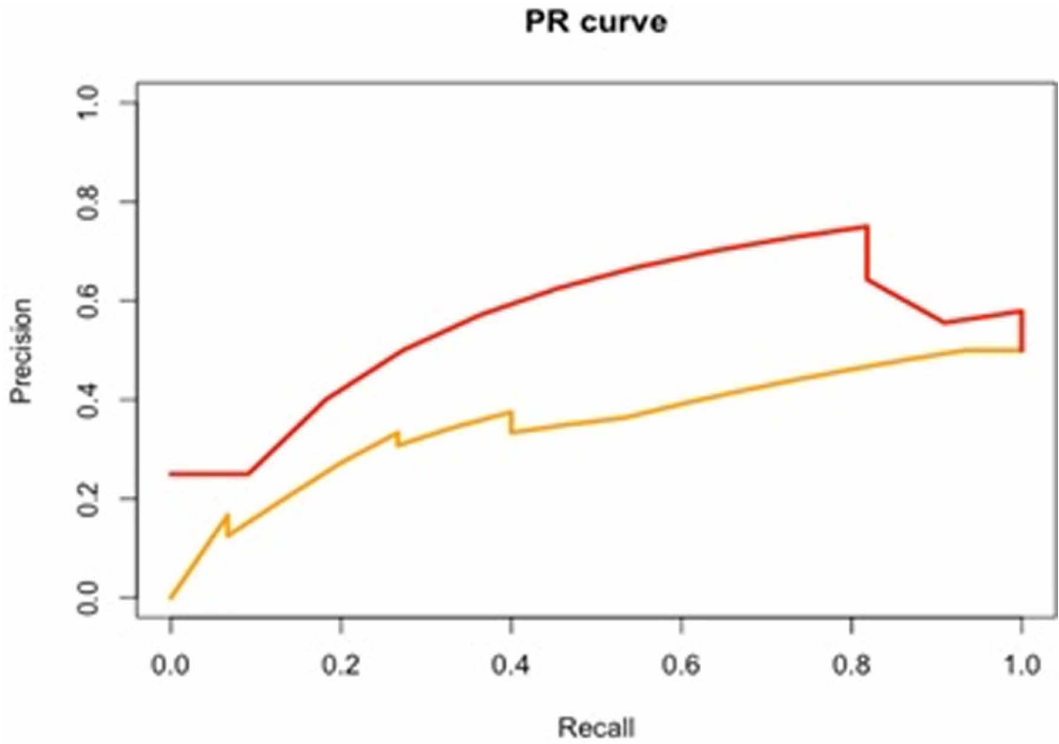

Figure 10. Overlapping ROC curves: red for A-priori, green for Utility based algorithm
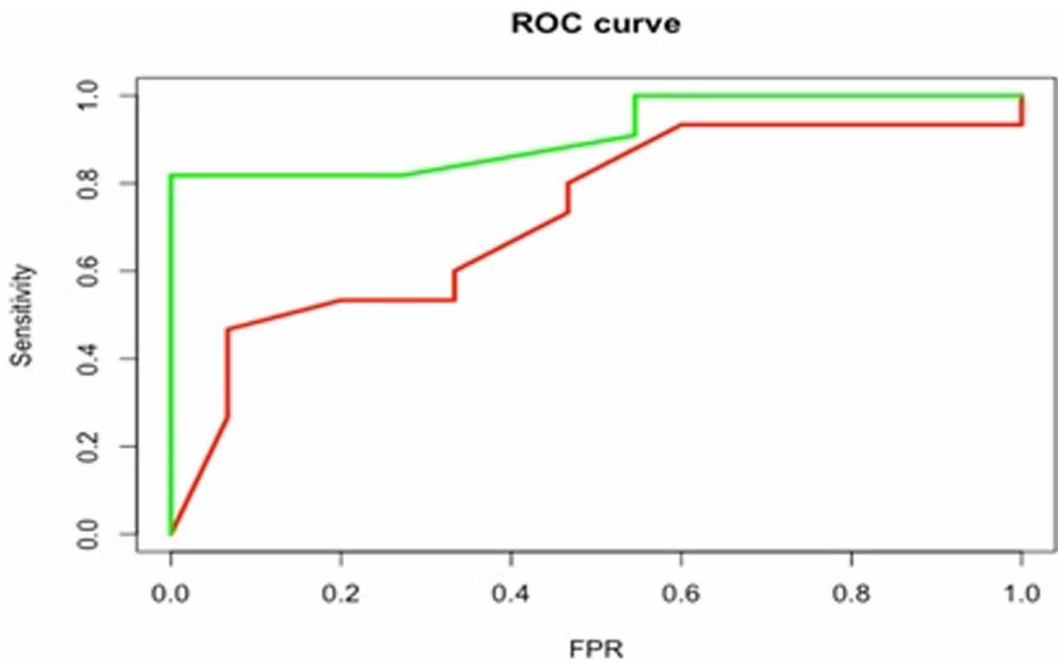
Table 3. Comparison Measures of Utility based Association rules and A-priori algorithm

Comparison Measures	A-priori Algorithm	Improved Utility Based Algorithm
Precision	0.54	**0.61**
Recall	0.78	**0.94**
F-Score	0.63	**0.75**
AUC -ROC	0.72	**0.91**

*Bold values represent better performance

- The recommender system based on the improved utility based algorithm have achieved a percentage improvement of 12% over the recommender system based on A-priori algorithm.
- Similarly, for the recall measure an improvement of 20.5% is achieved.
- For F-score measure, the traditional algorithm is out-performed by the Utility based algorithm by 19%.
- The higher AUC for PR curve and ROC curve for utility based rules signifies a greater accuracy of the recommender system as opposed to association rule based system.

Summarizing, we can say that in the recommendation engine, we implemented all the rules we retrieved in order to check the various comparison measures between the prevalent traditional method i.e. the A-priori algorithm and our proposed method, the Utility based algorithm. We also incorporated a wholesome E-commerce website experience by providing various facilities such as User Login, like, comment and review faculties along with providing recommendations while the items are in cart and also when they've been checked out. However, due to the high complexity of algorithms running time was high causing the process of rule retrieval a time taking one. Also the threshold value to be set for the algorithms needed to defined carefully so that the result contained all the rules that could of value to us. A very high threshold value in filtering could let some valuable results get lost and a very low value would cause rules that are unnecessary also creep into the final result.

CONCLUSION AND FUTURE WORK

Initially introduced by the means of A-priori algorithm, one of the most popular algorithm for association rule mining from transactions given. Having satiated the minimum support and confidence requirements as per the users' need, Frequent Item Set Mining algorithm however estimates more than actually desired. Therefore, ever since then efforts were made to retrieve useful rules from the databases and various efficient methods have been proposed. In the current scenario, Utility Mining made a place for its own and is one of the highly researched fields. Because of the added possibility of considering utility of item for increased the whole algorithm has gained a lot of momentum in the field of business. Therefor we in this paper tried to recreate a method in such a manner that not only the conventional method approach but even the current requirements are incorporated in one big picture. The implementation of utility-based method on a recommender system gave some outstanding results.

For future study, an analysis helping in yielding results faster i.e. better time complexity, and better space complexity can be considered. With the growing technological aspects an algorithm as credible and efficient as the conventional but as useful as Utility Based ones will only expand the world's outlook on Market Basket Analysis and its applications. The increasing customer base and the competitive market environment from a data so useful, which can help only in expanding the horizon of not only for service providers but also for service receivers. Recommendations not only work out well for e-commerce websites or movies but is today a part of almost all platforms available.

It is rightly said many times people do not know what they want until they see it. Recommendation engines in the coming time will work to offer a more immersive, visceral and well-rounded experience enhancing a customer's spree all together.

REFERENCES

Abraham, M. M., & Lodish, L. M. (1993). An implemented system for improving promotion productivity using store scanner data. *Marketing Science*, *12*(3), 248–269. doi:10.1287/mksc.12.3.248

Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. *SIGMOD Record*, *22*(2), 207–216. doi:10.1145/170036.170072

Bayardo, R. J. Jr, & Agrawal, R. (1999, August). Mining the most interesting rules. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 145-154). ACM. doi:10.1145/312129.312219

Bhattacharya, S., & Dubey, D. (2012). High Utility Itemset Mining. *International Journal of Emerging Technology and Advanced Engineering*, *2*(8), 476–481.

Buono, P., & Costabile, M. F. (2005). Visualizing Association Rules in a Framework for Visual Data Mining. From Integrated Publication and Information Systems to Virtual Information and Knowledge Environments, 221–231. doi:10.1007/978-3-540-31842-2_22

Burke, R. R. (1996). Virtual shopping: Breakthrough in marketing research. *Journal of Product Innovation Management*, *6*(13), 558–559.

Cakir, O., & Aras, M. E. (2012). A recommendation engine by using association rules. *Procedia: Social and Behavioral Sciences*, *62*, 452–456. doi:10.1016/j.sbspro.2012.09.074

Chen, Sain, & Guo. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. Journal of Database Marketing and Customer Strategy Management, 19(3), 197-208. 10.1057/dbm.2012.17

Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H. (2005). Market basket analysis in a multiple store environment. *Decision Support Systems*, *40*(2), 339–354. doi:10.1016/j.dss.2004.04.009

Cheng, J., Sun, A. R., Hu, D., & Zeng, D. D. (2010). *An information diffusion based recommendation framework for micro-blogging*. Academic Press.

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., & Sampath, D. et al. (2010, September). The YouTube video recommendation system. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 293-296). ACM. doi:10.1145/1864708.1864770

Ertek, G., & Demiriz, A. (2006). A Framework for Visualizing Association Mining Results. In *ISCIS* (pp. 593–602). doi:10.1007/11902140_63

Gatzioura, A., & Sànchez-Marrè, M. (2015). A case-based recommendation approach for market basket data. *IEEE Intelligent Systems*, *30*(1), 20–27. doi:10.1109/MIS.2014.57

Guy, I., Zwerdling, N., Ronen, I., Carmel, D., & Uziel, E. (2010, July). Social media recommendation based on people and tags. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 194-201). ACM. doi:10.1145/1835449.1835484

Hahsler, M., & Chelluboina, S. (2011). Visualizing association rules: Introduction to the R-extension package arulesViz. *R Project Module,* 223-238.

Hallowell, R. (1996). The relationships of customer satisfaction, customer loyalty, and profitability: An empirical study. *International Journal of Service Industry Management*, *7*(4), 27–42. doi:10.1108/09564239610129931

Han, J., Pei, J., & Yin, Y. (2000, May). Mining frequent patterns without candidate generation. *SIGMOD Record*, *29*(2), 1–12. doi:10.1145/335191.335372

Hernando, A., Bobadilla, J., & Serradilla, F. (2010). Collaborative Filtering Based on Choosing a Different Number of Neighbors for Each User. In *Handbook of Social Network Technologies and Applications* (pp. 317–330). Springer. doi:10.1007/978-1-4419-7142-5_15

Hu, J., & Zhang, B. (2012). *Product Recommendation System*. CS224W Project Report.

Kim, H., Kim, J., & Chen, Q. (2012). A product network analysis for extending the market basket analysis. *Expert Systems with Applications*, *39*(8), 7403–7410. doi:10.1016/j.eswa.2012.01.066

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., & Verkamo, A. I. (1994). Finding Interesting Rules from Large Sets of Discovered Association Rules. *CIKM*, 401–407. https://doi.acm.org/10.1145/191246.191314

Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2001, November). Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd international workshop on Web information and data management* (pp. 9-15). ACM. doi:10.1145/502932.502935

Mostafa, M. M. (2015). Knowledge discovery of hidden consumer purchase behaviour: A market basket analysis. *International Journal of Data Analysis Techniques and Strategies*, *7*(4), 384–405. doi:10.1504/IJDATS.2015.073867

Özseyhan, C., Badur, B., & Darcan, O. N. (2012). An association rule-based recommendation engine for an online dating site. *Communications of the IBIMA*, *2012*, 1. doi:10.5171/2012.775316

Rainsford, C. P., & Roddick, J. F. (2000). Visualisation of Temporal Interval Association Rules. In *IDEAL '00: Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning, Data Mining, Financial Engineering, and Intelligent Agents* (pp. 91–96). Springer-Verlag.

Russell, G. J., & Petersen, A. (2000). Analysis of cross category dependence in market basket selection. *Journal of Retailing*, *76*(3), 367–392. doi:10.1016/S0022-4359(00)00030-0

Seno, M., & Karypis, G. (2005). Finding frequent patterns using length-decreasing support constraints. *Data Mining and Knowledge Discovery*, *10*(3), 197–228. doi:10.1007/s10618-005-0364-0

Srikant, R., Vu, Q., & Agrawal, R. (1997, August). Mining association rules with item constraints. In Kdd (Vol. 97, pp. 67-73). Academic Press.

Srivastava, N., Gupta, K., & Baliyan, N. (2018). *Improved Market Basket Analysis with Utility Mining*. Academic Press.

Wong, P. C., Whitney, P., & Thomas, J. (1999). Visualizing Association Rules for Text Mining. In *INFOVIS '99: Proceedings of the 1999 IEEE Symposium on Information Visualization* (p. 120). IEEE Computer Society. doi:10.1109/INFVIS.1999.801866

Yang, L. (2003). Visualizing Frequent Itemsets, Association Rules, and Sequential Patterns in Parallel Coordinates. *Computational Science and Its Applications – ICCSA 2003*. *Lecture Notes in Computer Science*, *2667*, 21–30. doi:10.1007/3-540-44839-X_3

Ankita Verma is currently working as Assistant Professor in Jaypee Institute of Information Technology. She has acquired her PhD degree from Jawaharlal Nehru University, New Delhi. Her research interests include social network analysis, computational web intelligence, and machine learning.