

Data Mining Framework for Generating Sales Decision Making Information Using Association Rules

Md. Humayun Kabir

Dept. of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka-1342, Bangladesh

Abstract—The rapid technological development in the field of information and communication technology (ICT) has enabled the databases of super shops to be organized under a countrywide sales decision making network to develop intelligent business systems by generating enriched business policies. This paper presents a data mining framework for generating sales decision making information from sales data using association rules generated from valid user input item set with respect to the sales data under analysis. The proposed framework includes super shop's raw database storing sales data collected through sales application systems at different Point of Sale (POS) terminals. Apriori algorithm is famous for association rule discovery from the transactional database. The proposed technique using customized association rule generation and analysis checks the input items with sales data for validation of the input items. The support and confidence of each rule are computed. Sales decision making information about input items is generated by analyzing each of the generated association rules, which can be used to improve sales decision making policy to attract customers in order to increase sales. It is hoped that this approach for generating sales decision making information by analyzing sales data using association rules is more specific decision and application oriented as the business decision makers are not usually interested to all of the items of the sales database for making a specific sales decision.

Keywords—databases; data mining framework; Apriori algorithm; association rule; sales decision making information

I. INTRODUCTION

With the huge growing size of data and information in our modern technology world, organizational decisions are largely becoming dependent on computerized systems. Various sectors, e.g., education, agriculture, garments, stock exchange, finance and banking, super shops and many other sectors have potential applications of data mining systems in generating business intelligence (BI) for effective operations and efficient decision making. In recent days, massive data are collected from the customer's purchase records using application software system in the super shops, which are stored in databases using relational database management systems (RDBMSs), e.g., Oracle, MySQL.

Interesting patterns can be discovered from the customer's purchase records by analyzing the super shop's sales database

This research work is based on the work presented in [1] which was funded by the University Grants Commission (UGC) of Bangladesh in 2011-2012 in the Faculty of Mathematical and Physical Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh.

using the data mining techniques, which may reflect customer's buying patterns [1] [2]. The business authority may use this buying pattern to predict customer's buying habit and frequency of buying a particular item in the super shop. This information can be used productively by the business decision makers for efficient stock management and customer attraction to promote sales [1].

In data mining, intelligent algorithms are applied to organizational operational data stored in RDBMS or consolidated historic data stored in data warehouses (DWs) to extract hidden knowledge, interesting patterns, missing values, and new rules to speed up organizational decision making process [1]. Characterization, classification, association rule discovery and analysis, and clustering are the major data mining tasks, which are now widely used in various organizational applications. Each type of tasks has one or more algorithms which have been developed to perform data mining on transactional databases. The algorithm for mining association rules has been introduced in [3] in 1993. Apriori [4] is the famous algorithm for discovering association rules from the transactional database. Various research works [5] [6] [7] have been performed on this algorithm in various application domains. In association rule mining [4] [5] [6] [7] [8], association rules are generated by discovering relationship between the items of the transactions stored in the operational databases. Millions of sales transactions are performed in the supermarkets which are stored in operational databases for future processing. The management information system (MIS) and decision support system (DSS) personnel of a business organization must have the information about the customers buying habits, the best selling items, the items which make the highest profit, and the items not sold at all to develop attractive business policies to improve customer service and to increase sales. Infrequent items can be placed on offer based on the sales decision making knowledge extracted by analyzing sales data for each of the generated association rules to avoid business loss [1]. Association rule analysis [6] [7] [8] can be applied to the sales records of the supermarkets or sales centers to discover the buying patterns of customers. Based on this buying pattern, certain items can be placed on offer on discount to attract customers of a particular buying habit to buy a particular item which he was not used to buying. It is possible to generate association rules using only input item set rather than all items contained in the transactional database, so

that the generated rules can be useful in business decision making to promote sales on the specific item(s).

In business intelligence [9] [10], data of organizations is transformed into information to extract knowledge for effective decision making. Some recent research works on association rules are presented in [11] [12] [13]. Efficient algorithms have been introduced in [11] to generate candidate itemsets from transactional databases. These algorithms reduce candidates and also improve runtime for long transactions. An efficient approach for mining useful association rules from large transactional database using a clustering method to classify database and soft sets have been presented in [12]. A new association rule mining algorithm CMARM based on Confabulation-MapReduce is presented in [13] for analyzing medical data. This algorithm is also useful for infrequent items. In this paper, an approach for discovering business intelligence from the transactional database is presented for data mining using association rule generation and analysis technique [1]. Association rules generated from user input item set can be analyzed using sales data to generate information about sales items which can be used to improve sales decision making process. This approach first verifies whether the user input items exist within the transactional database, and then generates association rules from the valid input items only. Next, each of these generated association rules is analyzed based on the operational data records stored in the sales relation of the transactional database. The support and confidence [4] [5] [6] [7] of each rule is computed. A table of the generated association rules with their support and confidence is provided in each analysis case, which can be used to further analyze sales data to generate sales decision making information to obtain improved sales policy.

The remainder of this paper is organized as follows. Section II describes how data mining technique can be employed in the super shop's sales system using association rules. Section III presents a framework for data mining from relational databases as well as data warehouses. Section IV presents experimentation and results to demonstrate how the customized association rule generation and analysis technique can be used to generate sales decision making information. Finally, Section V concludes with a guideline to the future work.

II. APPLICATION OF DATA MINING TECHNIQUE IN SALES SYSTEM USING ASSOCIATION RULES

Today's business organizations are facing challenges on how they will attract customers. They have to change their business policies by adopting information and communication technology (ICT) incorporating intelligent software system within their automated business management system to speed up the business activities, quality of services and efficient policy making. Millions and millions of sales data records are collected through daily product sales at supermarkets and business organizations [1] [2], which are stored in relational databases using RDBMSs. This massive customer data may reflect customer's buying habits and frequency of buying a particular item.

Data mining can be performed on data records stored in a single relation or multiple relations. A super shop database may contain some relations containing data about the super shop, item details, supplier details, purchase, sales, members, sales man, customer, gifts, bill etc. [2]. In the scope of this paper, the purpose is to extract knowledge about customer's buying pattern from a sales database using association rules, and hence, a single relation called *itemssold* is used for a sales system. In this section, we consider this relation with the following attributes which store super shop's sales data of some customers.

itemssold(salesdate, item1, item2, item3, ..., itemn)

An association rule [3] [4] [5] [6] [7] [8] is a form of expressing the relationship between two data item sets. Association rules can be generated from items representing daily sales records stored in a super shop's operational database, which can be used to identify frequently purchased item sets of the customers [1]. The rule discovery usually considers data items stored in a single relation. The information about data items contained in the operational database can be expressed by the frequency of occurrence of the data items of a rule by measuring support and confidence of each rule. If X and Y are two sets of items, and I is the set of items in a super shop's sales transaction, then an association rule can be expressed in the form $X \Rightarrow Y$ [3] [4] [5] [6] [7] where $Y = I \setminus X$, $X \cap Y = \{\}$, i.e., empty set, X is the rule antecedent, Y is the rule consequent, X and Y are non-empty subsets of I , and \Rightarrow is an implication operator. TABLE I shows some customer data records stored in the *itemssold* relation of a transactional database.

TABLE I. CUSTOMER'S PURCHASE RECORDS STORED IN ITEMSSOLD RELATION [1]

Date	Items				
Sales date	item1	item2	Item3	item4	item5
04/09/2012	Beef	Ruhi	Fine Rice	Milk	Soft Drink
20/10/2012	Onion	Hilsha	Carrots	Bread	Milk
10/11/2012	Beef	Hilsha	Fine Rice	Milk	Soft Drink
25/11/2012	Onion	Rice	Banana	Bread	Dal
05/12/2012	Beef	Shrimp	Fine Rice	Milk	Cooking Oil

An association rule $r1$ can be defined using the items {Beef, Fine Rice, Milk, Soft Drink} shown in TABLE I in the following form [1]:

{Beef, Fine Rice, Milk} \Rightarrow {Soft Drink}

The support s and confidence c of an association rule $X \Rightarrow Y$ can be defined in terms of the number of occurrences of the item set in the antecedent X and consequent Y of a rule within a database relation containing N transactions. The support and confidence of an association rule are defined in [3] [4] [5] [6] [7] [8]. It can be computed as follows [4] [5] [6] [7] [8].

Support $s(X \Rightarrow Y)$

$$= ((\text{Number of records containing items } X \cup Y) / N) \times 100$$

The support s is 40% for rule $r1$. In this case, 2 records contain the items of $r1$ and $N = 5$ for the transaction records shown in TABLE I.

Confidence $c(X \Rightarrow Y)$

$$= ((\text{support of } X \cup Y) / \text{support of } X) \times 100$$

where

support of $X \cup Y$

$$= ((\text{Number of records containing items } X \cup Y) / N) \times 100$$

support of X

$$= (\text{Number of records containing items } X / N) \times 100$$

The confidence c of rule $r1$ is 66.66% for the transaction records shown in TABLE I.

III. SYSTEM ARCHITECTURE FOR SALES DATA MINING FRAMEWORK

The system architecture of a sales data mining framework for the application of data mining technique to the sales database of a typical super shop is depicted in Fig. 1. The framework includes two data storage components for mining knowledge: one is relational database component which stores daily sales data of the super shop, and the other is a data warehouse (DW) component which is constructed from the operational databases.

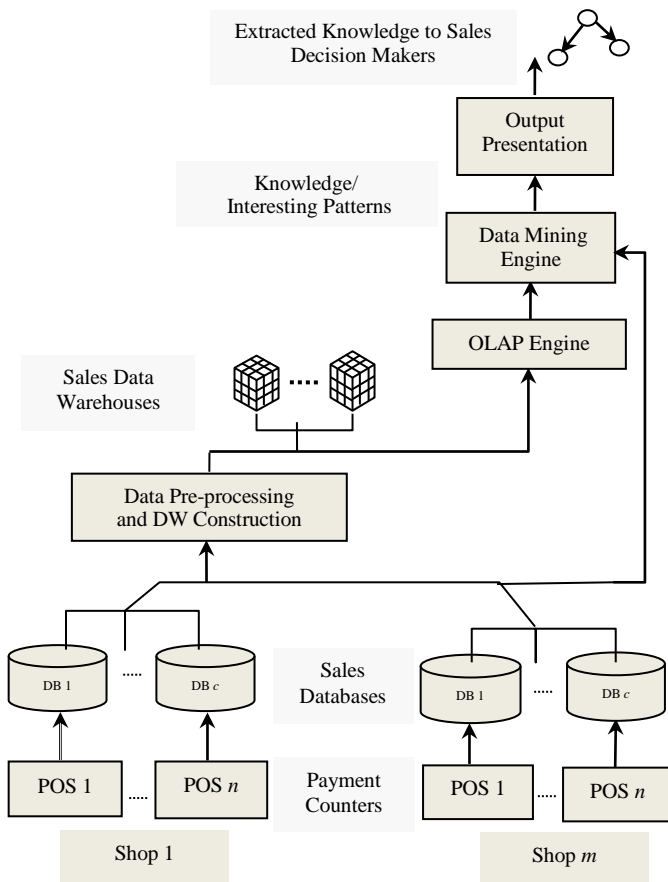


Fig. 1. System Architecture of a Sales Data Mining Framework

The point of sale (POS) terminals of the super shops are used to collect daily sales data through sales application software system, and the sales data is stored in the relational databases DB_1, \dots, DB_c maintained by the database servers. Data mining algorithms for classification, association rule analysis, and clustering can be employed in the data mining engine as required by the sales system to extract sales decision making knowledge. The framework shows database(s) which may consist of multiple relations, and in that case, multi-relational data mining techniques may need to be employed. For mining knowledge from DWs, required data may be selected from the DW(s) using OLAP engine. Each of the input data sources is treated differently by the mining modules within the data mining engine, as the DWs may be constructed using different schema structures [8] and may have different data content than that of the relational databases. The mined output knowledge is represented graphically to the sales decision makers through the output presentation component to be used in efficient decision making. It is a 7 layer architecture for knowledge discovery from databases. Among the layers, only 4 layers are used in data mining from sales databases whereas all of the 7 layers are required for data mining from sales data warehouses.

IV. EXPERIMENTATION AND RESULTS

The performance of association rule generation technique and the application of these association rules in mining knowledge from sales database for using in sales decision making are experimented in two different approaches which are explained below.

A. Approach 1

In this approach, Weka 3.4.3 Associator¹ and Weka 3.7.12 Associator² are applied to the purchase records of some customers shown in TABLE II after removing the *salesdate* attribute from the *itemssold* relation to discover association rules using Apriori algorithm.

TABLE II. PURCHASE RECORDS [1]

Items				
item1	item2	item3	item4	item5
Beef	Ruhi	Fine rice	Milk	Soft drink
Onion	Hilsha	Carrots	Bread	Milk
Beef	Hilsha	Fine rice	Milk	Soft drink
Onion	Rice	Banana	Bread	Dal
Beef	Shrimp	Fine rice	Milk	Cooking oil

The best association rules obtained by Weka 3.4.3 Associator using the customer's purchase records shown in TABLE II are listed below in TABLE III. In TABLE II, each record consists of a subset of the item set $I = \{\text{Beef, Ruhi, Fine rice, Milk, Soft drink, Onion, Hilsha, Carrots, Bread, Rice, Banana, Dal, Shrimp, Cooking oil}\}$ of 14 items in the set contained in the purchase records of customers. Only 3 items forming the subset $I = \{\text{Fine rice, Beef, Milk}\}$ out of these 14 items is used to generate 10 best rules as indicated in TABLE III using Weka 3.4.3 Associator. These rules can be classified into two categories: rules containing i) 2 items and ii) 3 items.

¹ url: <http://www.cs.waikato.ac.nz/ml/weka/>

² url: <http://www.cs.waikato.ac.nz/ml/weka/>

TABLE III. THE BEST RULES GENERATED USING WEKA 3.4.3 ASSOCIATOR [1]

Rule No.	Rule
1.	item3=Fine rice 3 ==> item1=Beef 3 conf: (1)
2.	item1=Beef 3 ==> item3=Fine rice 3 conf: (1)
3.	item4=Milk 3 ==> item1=Beef 3 conf: (1)
4.	item1=Beef 3 ==> item4=Milk 3 conf: (1)
5.	item4=Milk 3 ==> item3=Fine rice 3 conf: (1)
6.	item3=Fine rice 3 ==> item4=Milk 3 conf: (1)
7.	item3=Fine rice 3 item4=Milk 3 ==> item1=Beef 3 conf: (1)
8.	item1=Beef 3 item4=Milk 3 ==> item3=Fine rice 3 conf: (1)
9.	item1=Beef 3 item3=Fine rice 3 ==> item4=Milk 3 conf: (1)
10.	item4=Milk 3 ==> item1=Beef 3 item3=Fine rice 3 conf: (1)

The first 6 rules shown in TABLE III are generated using only any 2 items of the item subset l . Each of these 6 rules contains any two items of l . All of these two items of l are found in each of the 3 records out of the 5 purchase records shown in TABLE II. These 2-item rules contain items which have high occurrences (60%) within the purchase records. The remaining 4 rules contain all of the 3 items contained in the item subset l which are found in each of the 3 records out of the 5 purchase records shown in TABLE II. These 3-item rules contain items which also have high occurrences (60%) within the purchase records. The association rules which can be constructed using the items $l - l$ are ignored as these rules will contain low occurring items of the purchase records. The recent version of Weka 3.7.12 also produces the same association rules from the purchase records of TABLE II with some additional parameter values which are skipped here for the purpose of simplicity.

B. Approach 2

In this approach, customized association rule generation and analysis technique [1] is applied on user input item set to generate association rules which are then analyzed using the purchase records of customers to extract sales decision making information. In this analysis, all of the possible association rules which can be generated from the items of the input item set are only considered. The association rules generated from the input items not contained in the sales database are rejected. The association rule generation process starts by taking an item set as input, and generates the antecedent and consequent item sets taking all of the possible non-empty subsets of the input item set. In this customized approach, association rules are generated from item set generated only from the valid user input item set containing sales items rather than generating a full set of association rules containing all of the items or a subset of the items contained in the sales relation. This approach generates a small number of association rules for analyzing a query in relation to a specific business decision. In this case, the generated association rules are analyzed using a single sales relation called *itemssold* with the sales records shown in TABLE IV [1] to generate sales decision making information. In this approach, 4 analysis cases are performed for 4 different input item sets to generate sales decision making information from the generated association rules. In each analysis case, a table represents the number of the occurrences of the antecedent item set X and the consequent

item set Y within the sales relation by X_s and Y_s respectively along with the support s and the confidence c of each rule.

TABLE IV. PURCHASE RECORDS OF CUSTOMERS IN ITEMSSOLD RELATION [1]

ID	Items				
	item1	item2	item3	item4	Item5
0001	Beef	Ruhi	Fine rice	Milk	Soft drink
0002	Beef	Hilsha	Fine rice	Milk	Soft drink
0003	Beef	Hilsha	Fine rice	Milk	Soft drink
0004	Onion	Rice	Banana	Bread	Dal
0005	Beef	Hilsha	Fine rice	Milk	Soft drink
0006	Beef	Hilsha	Fine rice	Milk	Soft drink
0007	Onion	Hilsha	Carrots	Bread	Milk
0008	Beef	Fine rice	Hilsha	Milk	Soft drink
0009	Beef	Hilsha	Fine rice	Milk	Soft drink
0010	Beef	Shrimp	Fine rice	Milk	Cooking oil
0011	Beef	Ruhi	Fine rice	Milk	Soft drink
0012	Beef	Hilsha	Fine rice	Milk	Soft drink
0013	Beef	Hilsha	Carrots	Milk	Soft drink
0014	Onion	Rice	Banana	Bread	Fruit
0015	Beef	Shrimp	Fine rice	Milk	Cooking oil
0016	Beef	Katla	Fine rice	Milk	Pasta
0017	Onion	Hilsha	Carrots	Bread	Milk
0018	Beef	Hilsha	Fine rice	Milk	Fruit
0019	Beef	Hilsha	Fine rice	Milk	Soft drink
0020	Beef	Shrimp	Fine rice	Milk	Cooking oil
0021	Beef	Shrimp	Fine rice	Milk	Soft drink

The *customer* relation is used to inform sales offer information to the appropriate customer. In this case, the attributes of these two relations are considered as shown below.

customer(customerID, customerName, address, mobilePhone, email)

itemssold(customerID, item1, item2, item3, ..., itemn)

The customized association rule generation and analysis technique is used to analyze the customer's purchase records shown in TABLE IV for each of the generated association rules to generate sales decision making information. A number of analysis cases are completed for testing [14] [15] the customized association rule generation and analysis technique to generate sales decision making information represented with (customers, missing items) pair.

Definition 1: (*customers, missingitems*) pair. In this definition, for an association rule, a (*customers, missingitems*) pair is discovered where *customers* represent those customers who did not buy the items which are *missingitems* in the current purchase record for a successful matching of the rule antecedent X . Missing items in a matched record are those which are frequent items in other records, but missing in the currently matched record, so it may be placed on offer. The following symbols are used to determine a (*customers, missingitems*) pair.

$l_k \rightarrow$ rule item set

$l_a \rightarrow$ rule antecedent item set

$l_c \rightarrow$ rule consequent item set

$l_{k'} \rightarrow$ matched record item subset which is matched with a subset of the rule items

$l_r \rightarrow$ missing item set to be placed on offer
customers \rightarrow List of customers, initially {}
missingitems \rightarrow List of missing items, initially {}

If $match(l_k, l_r) = \text{False}$, then $(customers, missing\ items) = (\{\}, \{\})$. For matching to be successful, $l_k = l_k \cup l_r, l_r \subseteq l_k, l_k \subseteq l_r$. In the case of a successful matching, $(customers, missingitems) = (customers \cup customerID, missingitems \cup item_i)$ for $i = 1$ to n , where n is the number of items. For a particular association rule, the items of the missing item set to be placed on offer will be $l_r = l_k \setminus l_r$.

1) *Analysis Case 1:* The item set {Beef, Fine rice, Soft drink} is used as input to the customized association rule generation and analysis technique [1] to generate association rules consisting of all of these 3 input items. Each of these generated rules is analyzed using the customer’s purchase records shown in TABLE IV. The generated association rules with their corresponding support and confidence with the floating point values rounded up to 2 precision points are provided in TABLE V. Within the purchase records, only 11 records contain all of the rule items.

TABLE V. GENERATED ASSOCIATION RULES WITH THE SUPPORT AND CONFIDENCE USING CUSTOMIZED ASSOCIATION RULE GENERATION AND ANALYSIS TECHNIQUE [1]

Rule No.	Generated 3-item Association Rules	X_u	Y_u	s %	c %
1.	{Beef} \Rightarrow {Fine rice, Soft drink}	17	11	52.38	64.71
2.	{Fine rice} \Rightarrow {Beef, Soft drink}	16	12	52.38	68.75
3.	{Soft drink} \Rightarrow {Beef, Fine rice}	12	16	52.38	91.67
4.	{Beef, Fine rice} \Rightarrow {Soft drink}	16	12	52.38	68.75
5.	{Beef, Soft drink} \Rightarrow {Fine rice}	12	16	52.38	91.67
6.	{Fine rice, Soft drink} \Rightarrow {Beef}	11	17	52.38	100.0

The analysis of the generated association rules shown in TABLE V demonstrates that the rules consist of only input item set {Beef, Fine rice, Soft drink} among the 17 items $l = \{\text{Beef, Ruhi, Fine rice, Milk, Soft drink, Hilsha, Onion, Rice, Banana, Bread, Dal, Carrots, Shrimp, Cooking oil, Fruit, Katla, Pasta}\}$ contained in the *itemssold* relation shown in TABLE IV. Thus, only a small number of rules are analyzed to make any sales decision regarding these 3 input items forming a item group. This technique eliminates the need for analyzing a large number of rules that could be generated from all of the 17 items found within 21 data records shown in TABLE IV. The extraction of any sales decision making information usually relates to a single item, a pair or a group of 3 to 4 items. Customers often purchase items in pairs or groups, and the number of items in the group usually does not exceed 4. Some examples of such pairs of items or item groups which go together are {Computer, Software}, {Beef, Fine rice, Soft drink}, {Bread, Milk}, {Bread, Egg}, {Bread, Banana}, {Bread, Milk, Banana, Egg} etc. which may help to determine customers buying pattern. The missing of an item in a group within the customer’s purchase records may be considered as a data mining problem, and the frequency of occurrence of this item in the pair or item group within the

purchase records may be used in sales decision making. A large number of rules can be generated from the items of l taking a minimum of 2 items in the antecedent and consequent parts of a rule. Thus, the customized approach avoids the need for analyzing all of the association rules which can be generated from the 17 items. To make a sales decision in relation to any subset of the item set l , analyzing irrelevant rules may be cumbersome, time consuming and unrealistic also.

a) *Generated Sales Decision Making Information:* In analysis case 1, only 3-item rules shown in TABLE V are analyzed to generate sales decision making information by the customized association rule generation and analysis technique [1] with minimum support = 50% and minimum confidence = 60%. To assist in sales decision making, a list of customers is generated who may be interested in the corresponding items as listed in TABLE VI by analyzing each of the generated association rules shown in TABLE V using the purchase records of TABLE IV.

TABLE VI. GENERATED SALES DECISION MAKING INFORMATION FOR ITEM SET {BEEF, FINE RICE, SOFT DRINK} [1]

Rule No.	Customers	Items on Offer
1.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
1.	{0013}	{Fine rice}
2.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
3.	{0013}	{Fine rice}
4.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
5.	{0013}	{Fine rice}
6.	{}	{}

In the above analysis, it is assumed that the higher the support and the higher the confidence of an association rule over the minimum support and minimum confidence, it is more likely that the items within the rule under analysis would be more frequent in the customers buying pattern. Based on this assumption, the sales decision making information about customers and the items of interest, *i.e.*, missing item(s) in the purchase record(s), to be placed on offer is generated as shown in TABLE VI.

From the above analysis, it can be predicted that for any rule satisfying the minimum support and minimum confidence, the item(s) missing in a customer’s purchase record which contains a subset of the valid input items, all of the items of the rule antecedent X , and at least one item of the rule consequent Y , may attract some customers if the item(s) is placed on offer. From this analysis, we can also decide that the stock of these rule items should be increased.

2) *Analysis Case 2:* In this analysis case, the input item set {Beef, Fine rice, Milk, Soft drink} is used by the customized association rule generation and analysis technique [1] to generate 4-item association rules. Each of these generated rules is analyzed using the customer’s purchase records of *itemssold* relation shown in TABLE IV to generate sales decision making information. TABLE VII provides the generated association rules consisting of all of the 4 input items with their corresponding support and confidence with

the floating point values rounded up to 2 precision points. Within the purchase records, only 11 records contain all of the rule items.

TABLE VII. GENERATED ASSOCIATION RULES WITH THE SUPPORT AND CONFIDENCE USING CUSTOMIZED ASSOCIATION RULE GENERATION AND ANALYSIS TECHNIQUE

Rule No.	Generated 4-item Association Rules	χ_n	γ_n	s %	c %
1.	{Beef} \Rightarrow {Fine rice, Milk, Soft drink}	17	11	52.38	64.71
2.	{Fine rice} \Rightarrow {Beef, Milk, Soft drink}	16	12	52.38	68.75
3.	{Milk} \Rightarrow {Beef, Fine rice, Soft drink}	19	11	52.38	57.89
4.	{Soft drink} \Rightarrow {Beef, Fine rice, Milk}	12	16	52.38	91.67
5.	{Beef, Fine rice} \Rightarrow {Milk, Soft drink}	16	12	52.38	68.75
6.	{Beef, Milk} \Rightarrow {Fine rice, Soft drink}	17	11	52.38	64.71
7.	{Beef, Soft drink} \Rightarrow {Fine rice, Milk}	12	16	52.38	91.67
8.	{Fine rice, Milk} \Rightarrow {Beef, Soft drink}	16	12	52.38	68.75
9.	{Fine rice, Soft drink} \Rightarrow {Beef, Milk}	11	17	52.38	100.0
10.	{Milk, Soft drink} \Rightarrow {Beef, Fine rice}	12	16	52.38	91.67
11.	{Beef, Fine rice, Milk} \Rightarrow {Soft drink}	16	12	52.38	68.75
12.	{Fine rice, Milk, Soft drink} \Rightarrow {Beef}	11	17	52.38	100.0
13.	{Milk, Soft drink, Beef} \Rightarrow {Fine rice}	12	16	52.38	91.67
14.	{Soft drink, Beef, Fine rice} \Rightarrow {Milk}	11	19	52.38	100.0

a) *Generated Sales Decision Making Information:* In this analysis case, only 14 association rules consisting of 4 items shown in TABLE VII are analyzed by applying the customized association rule generation and analysis technique [1] for generating sales decision making information as provided in TABLE VIII considering minimum support = 50% and minimum confidence = 60%.

TABLE VIII. GENERATED SALES DECISION MAKING INFORMATION FOR ITEM SET {BEEF, FINE RICE, MILK, SOFT DRINK}

Rule No.	Customers	Items on Offer
1.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
1.	{0013}	{Fine rice}
2.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
4.	{0013}	{Fine rice}
5.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
6.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
6.	{0013}	{Fine rice}
7.	{0013}	{Fine rice}
8.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
9.	{}	{}
10.	{0013}	{Fine rice}
11.	{0010, 0015, 0016, 0018, 0020}	{Soft drink}
12.	{}	{}

13.	{0013}	{Fine rice}
14.	{}	{}

The sales decision making information about customers and the items of interest, i.e., missing item(s) in their purchase records, to be placed on offer is generated. In the above analysis, all of the association rules except rule 3 shown in TABLE VII satisfy the minimum support and minimum confidence. Rule 3 is excluded from sales decision making information generation as shown in TABLE VIII as it's confidence does not satisfy the minimum confidence level. Using the prediction of analysis case 1, it is expected that the items shown in TABLE VIII may attract some customers if the items are placed on offer. From this analysis, it may also be decided that the stock of these rule items should be increased. In TABLE VIII, the customer list and the items of interest to be placed on offer both are {} for the association rules 9, 12 and 14. In this generation process, the items contained in the antecedent X of the rule under analysis must be fully matched with any of the purchase record items, and also the rule consequent $Y \neq \{\}$ such that $\exists i \in Y$ and $i \in I$ where i is any item contained in the purchase records.

3) *Analysis Case 3:* The input item set {Onion, Hilsha, Bread, Milk} is used by the customized association rule generation and analysis technique [1] to generate 4-item association rules. Each of these generated rules is analyzed using the purchase records of the items sold relation to generate information for the valid rules only. TABLE IX provides the generated association rules consisting of all of the input items with their corresponding support and confidence with the floating point values rounded up to 2 precision points.

TABLE IX. GENERATED ASSOCIATION RULES WITH THE SUPPORT AND CONFIDENCE USING CUSTOMIZED ASSOCIATION RULE GENERATION AND ANALYSIS TECHNIQUE

Rule No.	Generated 4-item Association Rules	χ_n	γ_n	s %	c %
1.	{Onion} \Rightarrow {Hilsha, Bread, Milk}	4	2	9.52	50.0
2.	{Hilsha} \Rightarrow {Onion, Bread, Milk}	12	2	9.52	16.67
3.	{Bread} \Rightarrow {Onion, Hilsha, Milk}	4	2	9.52	50.0
4.	{Milk} \Rightarrow {Onion, Hilsha, Bread}	19	2	9.52	10.53
5.	{Onion, Hilsha} \Rightarrow {Bread, Milk}	2	2	9.52	100.0
6.	{Onion, Bread} \Rightarrow {Hilsha, Milk}	4	12	9.52	50.0
7.	{Onion, Milk} \Rightarrow {Hilsha, Bread}	2	2	9.52	100.0
8.	{Hilsha, Bread} \Rightarrow {Onion, Milk}	2	2	9.52	100.0
9.	{Hilsha, Milk} \Rightarrow {Onion, Bread}	12	4	9.52	16.67
10.	{Bread, Milk} \Rightarrow {Onion, Hilsha}	2	2	9.52	100.0
11.	{Onion, Hilsha, Bread} \Rightarrow {Milk}	2	19	9.52	100.0
12.	{Hilsha, Bread, Milk} \Rightarrow {Onion}	2	4	9.52	100.0
13.	{Bread, Milk, Onion} \Rightarrow {Hilsha}	2	12	9.52	100.0
14.	{Milk, Onion, Hilsha} \Rightarrow {Bread}	2	4	9.52	100.0

Within the purchase records, only 2 records contain all of the rule items. In this analysis case, only the generated 4-item association rules are analyzed for generating sales decision making information using the customized association rule generation and analysis technique [1] considering minimum support = 50% and minimum confidence = 60%. In this case,

no sales decision making information is generated as the support of each association rule is 9.52% which is less than the minimum support. We can identify the infrequent items from TABLE IX. The business authority must have to make efficient policies to increase the sales of these infrequent items and also make necessary steps for proper stock management.

4) *Analysis Case 4:* In this case, the input item set {Beef, Shrimp, Fine rice, Milk} is used by the customized association rule generation and analysis technique [1] to generate 4-item association rules as shown in TABLE X. Each of these generated association rules is analyzed using the customer's purchase records to generate sales decision making information. TABLE X provides the generated association rules with their corresponding support and confidence with the floating point values rounded up to 2 precision points. Within the purchase records, only 4 records contain all of the rule items. In this analysis, only 4-item rules are used for generating sales decision making information by the customized association rule generation and analysis technique [1]. Considering the minimum support = 50% and minimum confidence = 60%, no sales decision making information is generated as the support 19.05% of the generated rules does not satisfy the minimum support value.

TABLE X. GENERATED ASSOCIATION RULES WITH THE SUPPORT AND CONFIDENCE USING CUSTOMIZED ASSOCIATION RULE GENERATION AND ANALYSIS TECHNIQUE

Rule No.	Association Rules Consisting of 4 Input Items	X _n	Y _n	s %	c %
1.	{Beef} ⇒ {Shrimp, Fine rice, Milk}	17	4	19.05	23.53
2.	{Shrimp} ⇒ {Beef, Fine rice, Milk}	4	16	19.05	100.0
3.	{Fine rice} ⇒ {Beef, Shrimp, Milk}	16	4	19.05	25.0
4.	{Milk} ⇒ {Beef, Shrimp, Fine rice}	19	4	19.05	21.05
5.	{Beef, Shrimp} ⇒ {Fine rice, Milk}	4	16	19.05	100.0
6.	{Beef, Fine rice} ⇒ {Shrimp, Milk}	16	4	19.05	25.0
7.	{Beef, Milk} ⇒ {Shrimp, Fine rice}	17	4	19.05	23.53
8.	{Shrimp, Fine rice} ⇒ {Beef, Milk}	4	17	19.05	100.0
9.	{Shrimp, Milk} ⇒ {Beef, Fine rice}	4	16	19.05	100.0
10.	{Fine rice, Milk} ⇒ {Beef, Shrimp}	16	4	19.05	25.0
11.	{Beef, Shrimp, Fine rice} ⇒ {Milk}	4	19	19.05	100.0
12.	{Shrimp, Fine rice, Milk} ⇒ {Beef}	4	17	19.05	100.0
13.	{Fine rice, Milk, Beef} ⇒ {Shrimp}	16	4	19.05	25.0
14.	{Milk, Beef, Shrimp} ⇒ {Fine rice}	4	16	19.05	100.0

We can identify the infrequent items from TABLE X. The business authority must have to make efficient policies to increase the sales of these items and also make necessary steps for proper stock management.

V. CONCLUSION AND FUTURE WORK

In this paper, the system architecture of a data mining framework for the application of data mining technique to the sales database of a super shop has been presented. It is shown that association rules can be generated from valid input items with respect to the sales data under analysis contained in the sales database. The outcome of the application of the customized association rule generation and analysis technique

on sales database has been presented in this paper. The success of this technique in generating sales decision making information usually depends on the frequency of occurrences of the items within the sales data records, and the nature of customer's buying pattern of purchasing particular item groups which go together. The research work demonstrates how the generated association rules can be used to produce sales decision making information by analyzing sales data. Each of the generated association rules is verified to see whether the correct association rules are generated consisting of only user input items contained in the sales database. A table representing the number of occurrences of the rule antecedent, consequent, and the support and confidence of each rule for a given user input item set is provided. This work clarifies how the application of the data mining technique using association rules can help to improve decision making in a sales system using the generated sales decision making information. In this work, a single relation of sales data is used for generating sales decision making information by analyzing each of the generated association rules. To improve customer service using the generated sales decision making information, a customer database relation can be used to inform sales offer information to the appropriate customer. Future work may include multiple relations for association rule mining and analysis to generate sales decision making business intelligence. More research is needed for the analysis of less frequent items in sales systems. The improvement in the current implementation of the customized association rule generation and analysis technique may also be done in the near future.

ACKNOWLEDGMENT

The author would like to thank all of his colleagues and experts who have made valuable comments to this work. He also thanks all of the people who helped in any matter related to this research work.

REFERENCES

- [1] Md. Humayun Kabir, Integration and Testing of Association Rule Discovery Algorithm for Developing Masranga Data Mining Software System, Research Project Report, 2012, Faculty of Mathematical and Physical Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh.
- [2] Md. Humayun Kabir, Development of a Data Mining Software for Data Warehouse Construction and Mining Interesting Patterns from Organizational Databases, Research Project Report, 2010, Faculty of Mathematical and Physical Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh.
- [3] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Databases", Proceedings of the ACM SIGMOD Conference on Management of Data, Washington DC, pp. 207-216.
- [4] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules", Proceedings of the 20th VLDB Conference, 1994, Santiago, Chile, pp. 487-499.
- [5] R. Srikant and R. Agrawal, "Mining generalized association rules", Proceedings of the 21st VLDB Conference, 1995, Zurich, Switzerland, pp. 407-419.
- [6] M. Karaolis, J. A. Moutiris, L. Papconstantinou, and C. S. Pattichis, "Association Rule Analysis for the Assessment of the Risk of the Coronary Heart Events", in the Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, 2009 (EMBC 2009), pp. 6238-6241. IEEE Conference Publications.
- [7] Z. Zhu and J.-Y. Wang, "Book recommendation service by improved association rule mining algorithm", Proceedings of the 6th International

- Conference on Machine Learning and Cybernetics, Hong Kong, IEEE Conference Publications, 2007, pp 3864-3869.
- [8] S. Nestorov and N. Jukić, "Ad-Hoc association rule mining within the data warehouse", Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), IEEE Computer Society 2002.
- [9] A. Martin, T. M. Lakshmi, and V. P. Venkatesan, "A Business Intelligence Framework for Business Performance using Data Mining Techniques", in the International Conference on Emerging Trends in Science, Engineering and Technology 2012, IEEE publications, pp. 373-380.
- [10] R. T. Hans and E. MnKandla, "Modeling software engineering projects as a Business: A Business Intelligence Perspective", AFRICON 2013, IEEE Conference Publications, pp. 1-5.
- [11] V. S. Tseng, B.-E. Shie, C.-W. Wu and P. S. Yu, "Efficient Algorithms for Mining High Utility Itemsets from Transactional Databases", IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 8, August 2013, IEEE Computer Society, pp. 1772-1786.
- [12] B. Li, Z. Pei, K. Qin, "Association Rules Mining based on Clustering Analysis and Soft Sets", IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. 2015, IEEE Computer Society, pp. 675-681.
- [13] J. Gautam and N. Srivastava, "Analysis of medical domain using CMARM: Confabulation Mapreduce association rule mining algorithm for frequent and rare itemsets", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 11, 2015, pp. 224-228.
- [14] J. Collofello and K. Vehathiri, "An Environment for Training Computer Science Students on Software Testing", Proceedings of the 35th Annual Conference on Frontiers in Education FIE '05, 2005, pp. T3E-6.
- [15] Y. Labiche, "Integration testing object-oriented software systems: An experiment driven research approach", 24th Canadian Conference on Electrical and Computer Engineering (CCECE) 2011, IEEE Conference Publications, pp. 652-655.