



Analyze and Enhance Sales in Lulu Supermarket using Data Mining Technology

Ahmed Abdullah Awadh Koofan MIDDLE EAST COLLEGE

Mohammed Kaleem Middle East College

-Data mining is a powerful technology for analyzing huge data, it has many techniques such as; classification, clustering, prediction and association rules etc., In this research Association rule will be used for analyzing data, which will help to extract the data related to combinations of items. Numerous customers tends to purchase items regularly, each time they visit supermarket, customer's need to move around from shelf to shelf for the product of their interest which is time consuming. This research will help to minimize the time consumption for customers by analyzing the customer's invoices and letting know the supermarket about the patterns of customer's orientations. In this work python tool will be used for data mining, by using association rule to analyze the customer's purchases and retrieve the relevant information which will help to determine the customer's pattern and know the association between products. In this rationale, the data of customer's purchases were collected from Lulu hypermarket for data analysis and the outcomes of the analysis is to know the customer's patterns and making the shopping easy by reorganizing the related items and the most buying items together on same shelf.

Introduction

In this evolving world, most stores are now using new technology to records daily operations for further analysis. Data mining is one of the technique that is used often to extract the knowledge or behaviors from huge data which is important to the company [1]. Where the supermarkets are in great competition to preserve its benefits and win more customers, which encourage the market to use new tools to satisfy its customers and increase the company's revenues. Association rule is a model of data mining technique that analyzes the ordered dataset and gets more knowledge of the customer's interest in purchasing [2]. One of the most important aspects of data analysis is to facilitate customers for shopping by arranging items on shelves. Therefore, this research aims to use data mining techniques that sort data from huge lists and make it valuable to the company, which is one of the most used techniques to identify the pattern of customers and know the most important items that are always purchased with other items. This study focuses on using one of the data mining method, which is an association rule, this mechanism analyzes the data to determine the combination between purchases using Apriori algorithms.

Literature review

I. "Data Mining Applications for Sales Information System Using Market Basket Analysis on Stationery Company" [1]:

Data Mining is known as "write many of the reports and queries". Nonetheless, data mining not only query and report creation. It is completed with a specific tool, which performs data processing. Data mining is the mining of the patterns and the information that is important from the large database which some knowledge. Data mining study the data to extract some information from a



large volume of data. Data mining is a method that not the same statistical analysis. Tasks that are functional to data mining such as:

- Association: the procedure of finding the relations between items.
- Sequence: same association but in more than one period of time.
- Clustering: the process of collecting similar data to groups, each group has similar data.
- Classification: make a class of an object based on its concept.
- Regression: the method of approximating the value expected by the patterns based on the dataset.
- The solution: a procedure of finding a problem and explaining it to give some information to decision making.

Market basket analysis is an association in data mining to find elements that appear in one time. This method can know the customer buying patterns by finding an association between different items in different invoices. The results from these methods can be used by retail shopping such as sales company/supermarket to improve a marketing plan through items that may be purchased concurrently by consumers.

Examples of associations rule the possibility of how buyer to buy bread with milk every time. Association rules can be known for support and confidence. Support is a percentage of the combinations of items in the database, while confidence is a certain value of the strong association between items.

Association rules are regularly expressed in the form of: {bread, milk} → {butter} (support = 30% confidence = 65%). the rule means 65% of transactions includes bread, milk also contains butter. While 30% of all transactions having the butter.

Example of using the association rules in a supermarket is to know how the possibility for someone to buy bread together with butter. Therefore, the manager can change the location of the goods or redesigning the entire market, the benefit of support and confidence, the association rule can be divided into two parts; Analysis of the top regularity pattern and the founding of association rule.

Based on this article, there are many techniques for data mining, and the technique of association rule has been used to make better offers based on customer purchases. In this project, we will use the same technique to know the combination of items to provide the best offers and rearrange them to be easier for the customer.

II. “An efficient method for mining association rules” (cluster-based association rule (CBAR)) [3]:

The data mining tool defined as extracting useful knowledge from a large dataset, which can help the managers to create effective strategies, association rules are used to discover relationships and possible associations of elements or attributes between large data. These rules can be effective in detecting unknown relationships, and providing results can be the basis for solving the problem. They have proved to be very useful tools for the organization as it strives to improve its profitability and competitiveness.

The Apriori algorithm can detect meaningful groups and create correlation rules within large databases, but a large number of groups of candidate elements are created from individual groups.



This method also needs to check the entire database, level by level. So, they introducing a new algorithm called cluster-based association rule (CBAR), for effective mining of association rules, which overcomes the above. These methods focus on reducing the number of candidate groups, thus reducing the number of database checks. CBAR requires only one scan of the transaction database, This is not only limited to trimming large amounts of data, which reduces the time required for data scan and requires less variation, but also ensures the validity of results.

Based on this article, there is a solution that helps to accelerate the process of extracting data from a huge database through an algorithm that creates cluster tables by scanning the database once and then use the created tables to know the association between the goods sold. This technology is much faster than normal operation, which requires a lot of time in each reading the database.

III. “Business Data Mining – A Machine Learning Perspective” [6]:

Data mining is a machine learning method used to analyze data to discover patterns. To determine the appropriate method for an application, you must know the strengths and weaknesses of machine learning.

The method of finding a pattern of large data is not easy without the computer and the network, and using data mining technology can analyze the data from several sources easily than traditional statistics, as it depends on the following steps:

- Choose a dataset
- Clean selected data from inconsistencies
- Analyzing the dataset
- Evaluate the result (knowledge)

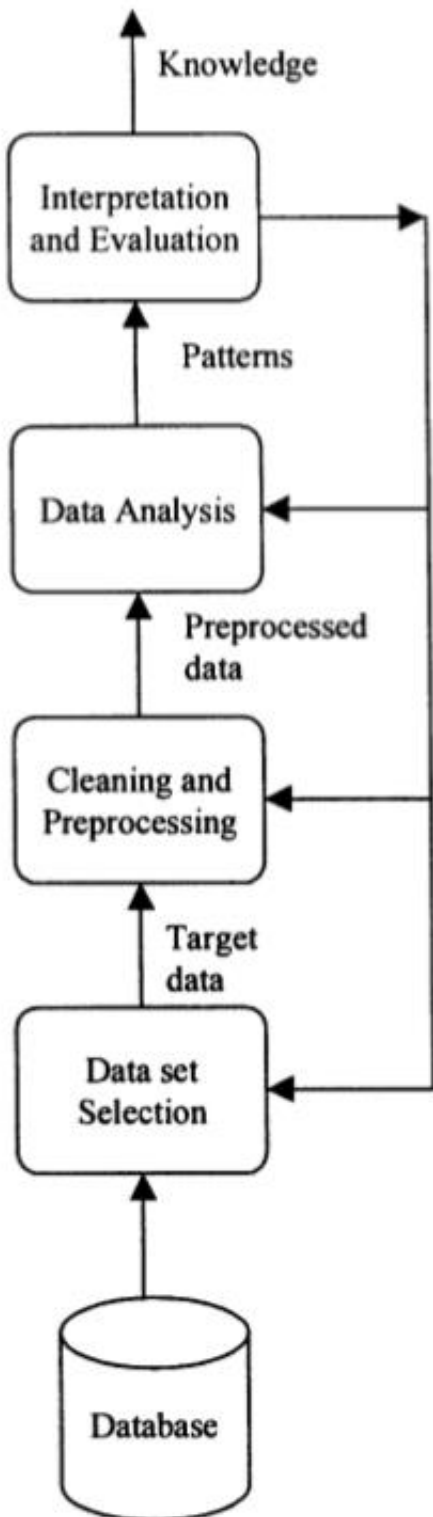


Figure 1. *Data mining process*

The article also classified applications in telecommunications, web, financial and market applications where the market side has several applications with different techniques in analyzing data to discover the association such as rule induction (RI) and visualization.

Found from the article that there are different techniques to get the association between sales,



such as market basket analysis application, which uses visualization technique which is the same technique that I will use in the project. Where the visualization technique is used to extract very useful information from a large dataset using some of these techniques: computer graphics, multimedia systems, human-computer interface, high-performance computing, and pattern recognition.

IV. “Efficient Association Rule Mining among both Frequent and Infrequent Items” [4]:

The primary purpose of using the technique of association rule in mining is to extract useful information from an item set. Also, the article describes that the association comes in the form of $A \Rightarrow B$ if A and B both are frequent items in the transactions which mean if the customer buys A almost buy B. So the items A and B can relocate close to each other or give a discount on these two items together which may increase the sales.

The article also discussed the traditional algorithm where it depends on the level of support and confidence when determining the level of support is very low resulting in poor performance while the determination of a high level of support resulted in the loss of many of the patterns that are rare in the transaction but an expensive. Moreover, the items that have high support can be easily identified, while those that have cross support may lack a correlation, and items that have low support often give a new insight.

In this article, I found two algorithms one is a matrix-based scheme (MBS) and the other is a hash-based scheme (HBS), these two algorithms are used for discovering the association rule from frequent and infrequent items but for frequent items it has some limitation on the length. These methods read the database two times only. Moreover, in the MBS use index function to know the items directly but it consumption large space in memory because it uses matrix so HBS comes to overcome the weakness of MBS. Therefore, I will not use one of these methods but I will use the Apriori algorithm.

V. Data Mining In Supermarket: A Survey [5]:

The most important motivation that made data mining always popular in the analysis of market data is to discover a new trend, the data mining process is for extracting the useful data that may benefit the market. The data mining technology is an operation to get useful patterns, information, and correlation from a set of data which helps in making a decision such as in the supermarket when analyzing the dataset and know the customer patterns or items correlations so the supermarket can select items that customers buy frequently to give a promote on it.

There are several techniques in data mining and it is divided into two groups descriptive or predictive where descriptive consist of association and clustering and the predictive consist of classification and prediction. The association rule in data mining is to get patterns from massive data, the form of an association rule is $A \rightarrow B$ which mean if the customer buys “A” expected to buy “B”. Where the support and confidence are the measures that use with association rule.

Support:

Confidence:

In this article, I get to know the four data mining techniques which are classification, prediction, association, and clustering. The classification is modeling the data to distinguish for a specific purpose. The prediction is to know what may happen in the future based on the available data. The association is to find frequent items in the transaction. The clustering is to determine the

relationship between objects that similar. From this discussion, I found that the association and clustering have some same properties but in the project, I focus on analyze frequent item which is in the association rule technique.

Methodology

In this research paper, the goal is to analyze the data of Lulu Super Market, which has many branches. So, we have taken the data of one branch to analyze it and get the customer's pattern in that branch using data mining. Where the working method is based on the CRISP-DM framework [9]. This framework has six phases showing in picture [1].

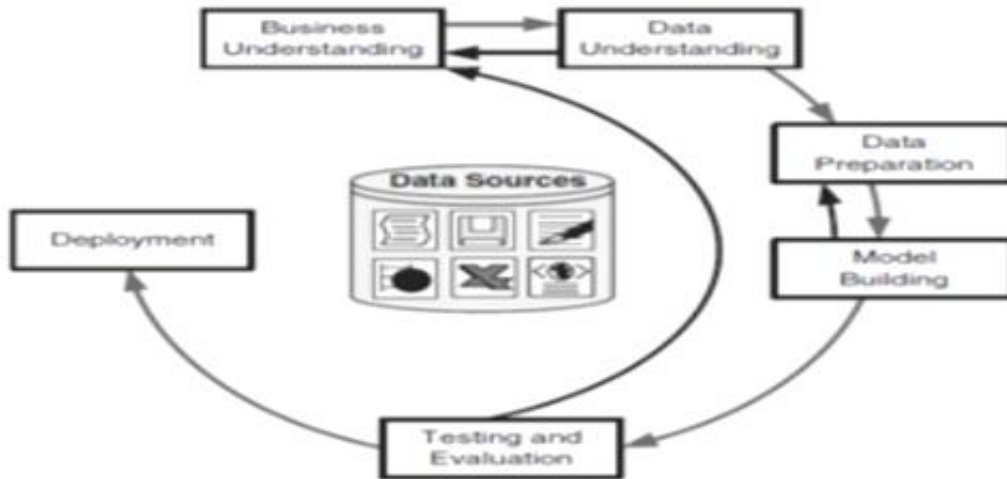


Figure 2. CRISP-DM [9]

First, understanding and identifying the business problem.

Second, determine the data set for analysis.

Third, prepare the dataset.

Fourth, build the models to calculate the value of support for each item.

Fifth, testing and analysis, to do the first sorting, set the minimum value of support then create a new data set of items that are higher than the minimum value, then set the combination of the two items rules for the remaining items, after that we calculate the support for the two items and the confidence and again set the minimum confidence value to create a last item set.

Sixth, deploying phase, eliminate unneeded rules.

Technology -Data mining

Data mining is one of the methods that use to analyzing the data to get useful knowledge, which can help the company to increase its revenue, it allows to analyze the data from different views, also classified the data, and identified the relationships [7]. Technically, the data mining process is to find the patterns from huge data, it sued on many sectors and the more used is in consumer retail that enables the companies to know the relationships between internal and external factors such as product, price, customer and the economic indicator.

Association rules



Association rule is a method of data mining which used to get interesting relationships between a large item-set. Now a day many researchers are interested in data mining -association rules- that mining the organization database to get the association rules. The knowledge association from huge customer purchasing orders can help in decision making such as marketing analysis and rearrange items position [7]. Our association rule method is shown in picture [1].

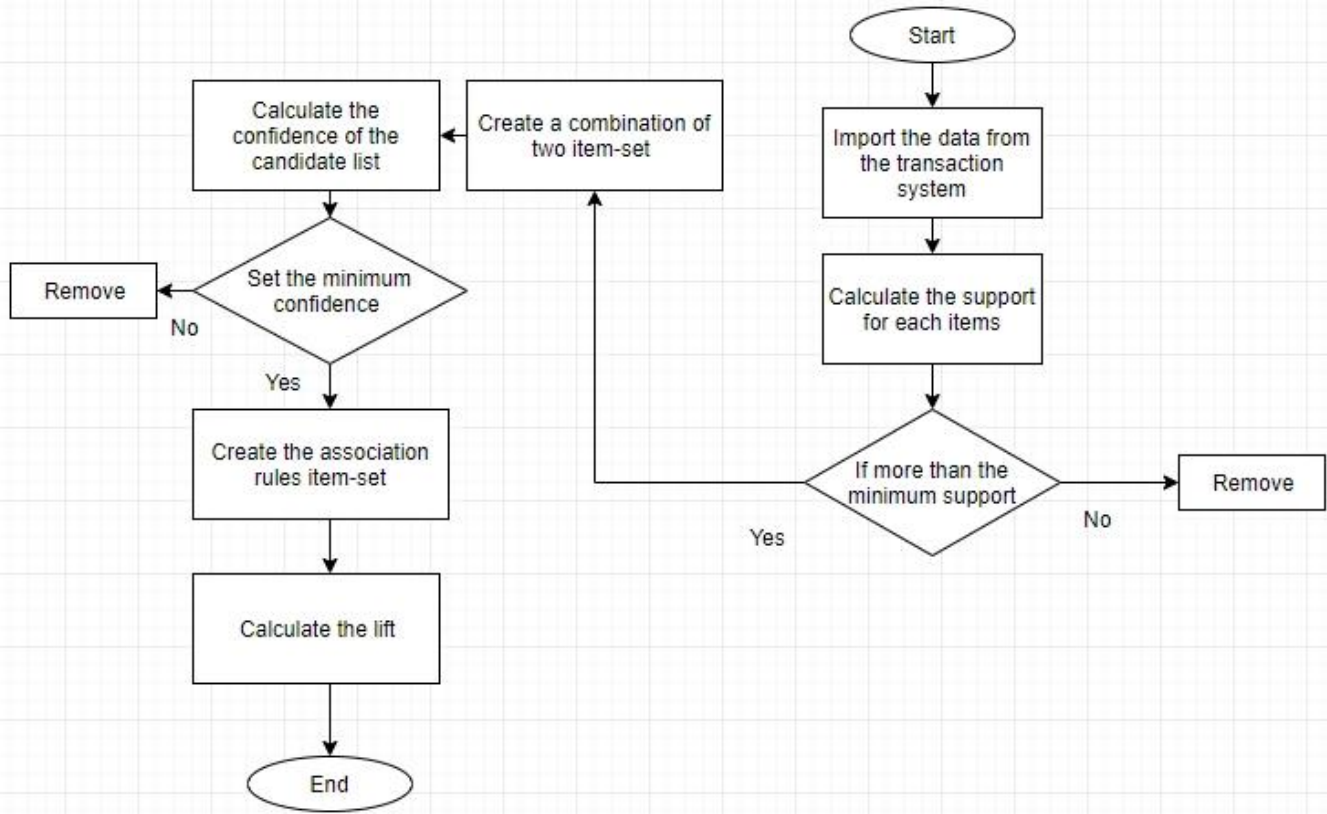


Figure 3. Flow chart

Apriori algorithms

The Apriori algorithm is the important algorithm used to mining the frequentative elements in a given dataset, which becomes a way to get the association rules. The useful of Apriori algorithm is using a breadth-first - searching level-wise - while searching, and k-things are to explaining the (k+1)-things. This algorithm uses bottom-up approach, which means the frequent subsets take one item on each process - generate a candidate group -, then test the groups of candidate items with the data. When there are no more extensions the algorithm will terminate. Then it removes infrequent items to generate k-length that has all frequent items [8].

Finding -result

This process is based on the analysis of purchasing orders from the transaction system. The purchasing items are shown in Table 1, from which we extract the support value for each item in the item set list shown in Table 2.

Order number	Purchasing items
1	A,D,E,C
2	F,D,A,C



3	A
4	D, B
5	A, B
6	F,D,A
7	E,F,C
8	A, D
9	A

Table 1. Purchasing items

Items Transaction	A	B	C	D	E	F
1	1	0	1	1	1	0
2	1	0	1	1	0	1
3	1	0	0	0	0	0
4	0	1	0	1	0	0
5	1	1	0	0	0	0
6	1	0	0	1	0	1
7	0	0	1	0	1	1
8	1	0	0	1	0	0
9	1	0	0	0	0	0
Total	7	1	3	5	2	2
Support	0.77	0.11	0.33	0.55	0.22	0.22

Table 2. Items classified with its support values

First, we extract the number of repetitions of items to calculate the value of support, which is the total transaction containing the item divided by total number of transactions [2].

Support rule [2]

From the values of support in the table [2], we can remove some items by minimum support, so here we can set our minimum support to 30% (0.30 or more). Therefore, the remaining items A, C, and D we have to give some rules as in table [3].

Item combination rules	support value	Confidence value
If A then C	0.22	0.28
If C then A	0.22	0.66
If C then D	0.22	0.66
If D then C	0.22	0.4
If D then A	0.44	0.8
If A then D	0.44	0.57

Table 3. calculating the confidence values

From the table [2] we can know the more sold item that has high support, but in our scenario we want to know the combination between items. Therefore, after calculating the support of a small combination of two items. Then we calculate the confidence which is the possibility of if the customer buys item A will also buy item C. Calculate the confidence is the support value of both items (A, C) divided by the support of the first item (A) [2].

Confidence rule [2]

From the table [3] the rule if D then A is the highest confidence which means the customer who buys D 80% will also buy A. so if we set the minimum confidence like 75% we are taking the high association rules.



Additionally, we also calculate the lift values to know the impact of the combination sales on the individual items sale [2].

Lift rule [2]

Item combination rules	support value	Confidence value	Lift value
If C then A	0.22	0.66	0.85
If C then D	0.22	0.66	2.18
If D then A	0.44	0.8	1
If A then D	0.44	0.57	1

Table 4. complete association rules

Finally, we got the strongest association rules from the item set. From the table [4] the best combination is {D -> A} and the high impact combination on the individual item sales is {C -> D}.

Future work

In this regard, data analysis is an important contribution to the knowledge of the direction of the business and always seek solutions to increase the profits in any trade. Also, this aspect also plays a big role in big data which is the trend of the era.

Conclusion

The data mining - association rules- techniques are very helpful for getting the buying pattern of customers in the supermarkets. With this research, we got that the final association rule has different habits of customer's buying. Based on these habits, the management can change its strategy to add different services to its customers to satisfy customers and win their loyalty to the commercial home.

Moreover, it is a useful system for the supermarket to know the different associations between products. Also, know the customer's interests and how he switches from a brand to another to meet his needs.

Reference

1. Setiawan, A., Satia Budhi, G., Haryadi Setiabudi, D., & Djunaidy, R. (2017). Data Mining Applications for Sales Information System by Using Market Basket Analysis on Stationery Company. *International Conference On Soft Computing, Intelligent System And Information Technology (ICSIIIT), Surabaya: Petra Christian University.*
2. Sagin, A., & Ayvaz, B. (2018). Determination of Association Rules with Market Basket Analysis: Application in the Retail Sector. *Southeast Europe Journal Of Soft Computing*, 7(1). doi: 10.21533/scjournal.v7i1.149
3. Tsay, Y., & Chiang, J. (2005). CBAR: an efficient method for mining association rules. *Knowledge-Based Systems*, 18(2-3), 99-105. doi: 10.1016/j.knosys.2004.04.010
4. Zhou, L., & Yau, S. (2007). Efficient association rule mining among both frequent and infrequent items. *Computers & Mathematics With Applications*, 54(6), 737-749. doi: 10.1016/j.camwa.2007.02.010
5. Kaur, D. and Kaur, J. (2017) "Data Mining In Supermarket: A Survey". *International Journal Of Computational Intelligence Research* 13 (8), 1945-1951
6. Bose, I., & Mahapatra, R. (2001). Business data mining — a machine learning perspective. *Information & Management*, 39(3), 211-225. doi: 10.1016/s0378-7206(01)00091-x
7. Raorane, A., & Kulkarni, R. (2011). Data Mining Techniques: A Source for Consumer



- Behavior Analysis. *International Journal Of Database Management Systems*, 3(3), 45-56. doi: 10.5121/ijdms.2011.3304
8. Abdulsalam, S., Adewole, K., Akintola, A., & Hambali, M. (2014). Data Mining in Market Basket Transaction: An Association Rule Mining Approach. *International Journal Of Applied Information Systems (IJ AIS)*, 7(10).
 9. Isa, N., Kamaruzzaman, N., Ramlan, M., Mohamed, N., & Puteh, M. (2018). Market Basket Analysis of Customer Buying Patterns at Corm Café. *International Journal Of Engineering And Technology*, 7(4), 119-123.