Journal of Theoretical and Applied Information Technology
15th December 2021. Vol.99. No 23
© 2021 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# IMPROVE SALES USING DATA MINING APRIORI ALGORITHM TO EXTRACT HIDDEN PATTERN

**[1] ABDALLA YASIN MABROUK, [2] MOHAMED HELMY KHAFAGY, [3]MOHAMAD ELMASRY, [4]MOHAMED HASAN IBRAHIM**

[1, 4] Department of Information System, Faculty of Computers and Information, Fayoum University, Fayoum, 63511, Egypt

[3] Department of Management Information Systems, Faculty of Business Administration, National Egyptian E-Learning University (EELU), Giza, 12611, Egypt

[2]Department of Computer Science, Faculty of Computers and Information, Fayoum University, Fayoum, 63511, Egypt

E-mail: [1]ay1112@fayoum.edu.eg, [2]mhk00@fayoum.edu.eg, [3] malmasry@eelu.edu.eg , [4]mhi11@fayoum.edu.eg

**ABSTRACT**

With the increasing growth of data in the world of business and companies, the amount of data resulting from sales operations in these companies increases, and there is no doubt that the presence of this data is a treasure. Through analyzing this data, unusual results can be reached that help decision-makers improve the profit process for these companies. We use in this research paper real data taken for the sales operations of Emisal company, which is located in Egypt and works on the sale of salts products during 2020 and four years back. Data mining technology is used in this paper, especially the apriori algorithm to explore the relationships between returns and the item, and the customer, and the month, and the day of the week, and the province. The results recommend eliminating the scenarios that may occur between the causes of return and the relationship with the item, customer, month, day of the week, and maintaining future sales.

**Keywords:** *Association Rules, Data Mining, Sales Analysis, Sales Returns, SPSS Modeler.*

## 1. INTRODUCTION

Data is around us from all sides. Data is a great resource that leads to improving an organization's business outcomes [1]. Data enables us to paint the most precise picture possible [2], Applying the best evidence and making informed decisions to promote any field. Data-driven decisions can remarkably improve services and build resilience to challenges [3].

There are many methods and sciences of data analysis to extract knowledge from these data [5]. In this study, the science of data mining is use [6][7]. Data mining is a process of extracting knowledge from data without any previous assumptions about what this knowledge could be to find patterns and correlations within large data sets [8].

In business, data analysis, especially corporate sales data, is indispensable for developing a corporate strategic plan to help reduce sales losses and increase profits for the company. [9][10], marketing activities, and advancements in manufacturing processes [11]. In addition, the analysis of the company's sales affects the future study of the market and competitors and knowledge of customer behavior, which facilitates the development of plans to attract customers. [12].

In this study, EMISAL Sales transaction data is used as a case study. EMISAL is a large industrial company located in the Arab Republic of Egypt. It produces salts after extracting them from Lake. It has a large amount of transactional sales data; this makes it difficult for decision-makers to analyze this data in the usual ways, so it was necessary to use modern techniques such as data mining to extract information that may be useful in improving the sales process. This study was conducted on actual data for four consecutive years, 2016 to 2020.

The main issue in this paper we try to solve, is that company EMISAL has a lot of data

due to it being a factory for salts products is a return of the product by the customer, and technically it is not possible to discover patterns in these data through primitive methods such as query in the database, it was necessary to use data mining to detect there are any relationships between a specific cause of reflux and a product, month, governorate, customer or factory.

The remaining sections of this paper are as follows: Literature review in the second section. Proposed methodology in the third section. Results in Section 4. & Discussion in section 5. Finally, the conclusion is giving in Section 6.

## 2. LITERATURE REVIEW

Abdalla Yasin Mabrouk, Mohamed Hasan Ibrahim, Mohamed Helmy Khafagy(2019), using data mining algorithms specifically association algorithms such as Apriori, Carma, and Sequence on real sales dataset taken from EMISAL company to make promotional in products that bought together with products that are not bought often to improve sales[13].

N. Isa, N.A.Kamaruzzaman M.A. Ramlan, N.Mohamed, M.Puteh, (2018) used data mining technology, especially the apriori algorithm, to analyze sales in stores that sell cosmetics, and the extent to which there are relationships between products and each other, and use these relationships in promoting marketing activities in the future and they found in the results of the analysis that there are some customers who may buy one product during a trip One shopping trip. On the other hand, some customers buy more than one product in one shopping trip. [14].

Setiawan, A., Budhi, G. S., Setiabudi, D. H., & Djunaidy, R (2017) used data mining market basket analysis method and administrative information systems-based websites to improve sales data control of business processes in the Stationery Company located in Indonesia. The results of this study indicate the relationships between products that are sold together [15].

Seren Sezen Karalök , Adnan Aktepe, Süleyman Ersöz (2016), using data mining algorithms specifically association algorithms such as GRI, Carma, and Apriori on real sales dataset taken from the large supermarket located in turkey to make market basket analysis to improve sales by extracting hidden pattern[16].

## 3. PROPOSED METHODOLOGY

In this study, Cross-Industry Standard Process for Data (CRISP-DM) [17][18], has been used to extract useful information and strong relation between returns sales and causes leading to return sales since the link between return sales and reasons are a many-to-many relationship.

### 2.1 CRISP – DM

As we mentioned previously that (CRISP-DM) has been used to extract useful information and strong relation between returns sales and causes leading to return sales since the link between return sales and reasons is a many-to-many relationship [19][20].

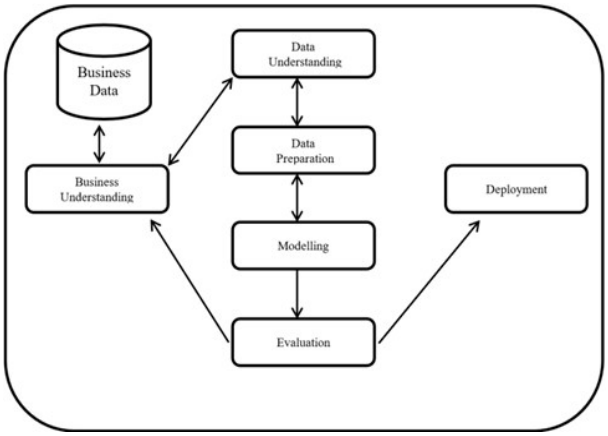


*Figure 1: Shows the CRISP-DM Cycle*

### 2.2 Business Understanding

In this phase, we will identify the contemporary issue faced by the EMISAL Company in Sales criteria. The interview session was conducted with the company's managers and staff, especially the sales and marketing sector boss. The interview session was conducted to understand and answer whether we wanted to be closer to the EMISAL Sales Process.

Before starting the process of analyzing any data, it is necessary to understand the sales cycle of the company, to facilitate understanding the sales data on the one hand, and to know the company's current situation on the other hand, which contributes to setting a future vision for the advancement of sales.

### 2.3 Data Understanding

As we mentioned previously, in this study, an interview was conducted for the specialized employees in the sales department of EMISAL and a sample was taken for sales from 2020 and four years back. The information that is collected describes the return of sales orders. Sixteen attributes have been extracted from the sales order: Return number, Sales order number, Plant ID, Date, Unit, Amount, Item ID, Item description, Customer ID, Customer name, Governorate, Type, Delivery type, Sales order, Reason ID, Reason Description. Table 1 below shows the metadata.

*Table 1: Shows the metadata of the data set.*

| Column Name | Data Type |
|---|---|
| Return number | int64 |
| Sales order number | int64 |
| Plant ID | int64 |
| Date | datetime64 |
| Unit | object |
| Amount | float64 |
| Item ID | int64 |
| Item description | object |
| Customer ID | int64 |
| Customer name | object |
| Governorate | object |
| Type | object |
| Delivery type | object |
| Sales order | int64 |
| Reason ID | object |
| Reason Description | object |

In this phase, before do mining in the data set, the datasets must be pre-processed from raw form. Since data are mostly not cleaned, we should pre-process data. The data pre-processing contains many steps that we follow in this study, the main steps contain consolidation, cleaning, transformation, and reduction of data.

In the process of collecting data for the study, and here we mean sales data, it is necessary to clean and standardize the form of the data, as this data was taken from more than one place in the company, and one of the most important ways to clean and standardize the form of data is to delete or complete any missing data, and reduce noise in the data, And delete any illogical data or affect the deviation from the normal values of the data.

In this study to work on raw data as inputs to the data mining algorithms, some columns were converted from their original shape to the desired shape and these columns are “Items”, “Plant”, “Customer”, “Governorate”. It should be noted that the date has extracted additional data, such as the month and days of the week. Figure 2 below shows the raw data before being transformed.

| | Sales Document | Plant | CustomerID | Customer_x | Item ID | Item | Amount | Unit |
|---|---|---|---|---|---|---|---|---|
| 1 | 10000011.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 2 | 10000008.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 200.000 | TO |
| 3 | 10000007.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 200.000 | TO |
| 4 | 10000005.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 5 | 10000003.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 6 | 10000016.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 7 | 10000014.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 8 | 10000021.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 9 | 10000020.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 10 | 10000034.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 11 | 27.000 | 4000.... | 100035.000 | الأستثمارات الصناعية | 200069.... | ملح كريات ماغنيسيوم مائية 25كغ ش منسوجة | 10.000 | TO |
| 12 | 10000041.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 13 | 10000037.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 14 | 29.000 | 4000.... | 100035.000 | الأستثمارات الصناعية | 200069.... | ملح كريات ماغنيسيوم مائية 25كغ ش منسوجة | 20.000 | TO |
| 15 | 10000036.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 16 | 10000032.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 17 | 10000047.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 18 | 33.000 | 4000.... | 100035.000 | الأستثمارات الصناعية | 200069.... | ملح كريات ماغنيسيوم مائية 25كغ ش منسوجة | 20.000 | TO |
| 19 | 10000052.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 20 | 10000049.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 21 | 10000058.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 22 | 10000056.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |
| 23 | 10000021.000 | 1000.... | 100035.000 | الأستثمارات الصناعية | 200000.... | ملح كريات صوديوم لامائية ج باج 1.25 طن | 100.000 | TO |

*Figure 2: Shows a sample of raw data in the SPSS Modeler.*

Table 2 below shows the shape of the data after processing it from its raw or original form to the form required in the applied algorithm. It contains a transaction ID, customer name, and reason to know if there is a relationship between the customer and specific reasons for return. Each reason contains a Boolean data type such as the "T" symbol, which means trustworthy or this customer has returned a product with this reason, and the "F" symbol, which means false or this customer has not given back a product with this reason.

*Table 2: Shows pre-process of data between the customer and returns reason.*

| Return Number | Customer Name | Return Reason Type 1 | Return Reason Type 2 | Return Reason Type 3 |
|---|---|---|---|---|
| 001 | Customer 1 | T | F | F |
| 002 | Customer 2 | F | F | T |
| … | … | … | … | … |
| n | Customer (n) | T | F | F |

Table 3 below shows the shape of the data after processing it from its raw or original form to the form required in the applied algorithm. It contains a transaction ID, Plant ID, and reason to know if there is a relationship between the plant and specific reasons for return. Each reason contains a Boolean data type with the "T" symbol, which means this plant has returned a product with this reason, and the "F" symbol, which means this plant has not produced a product with this reason.

*Table 3: Shows pre-process of data between plant and returns reason.*

| Return Number | Plant Name | Return Reason Type 1 | Return Reason Type 2 | Return Reason Type 3 | Return Reason Type N |
|---|---|---|---|---|---|
| 001 | Plant 1 | T | F | F | F |
| 002 | Plant 2 | F | F | T | F |
| … | … | … | … | … | … |
| n | Plant (n) | T | F | F | F |

Table 4 below shows the shape of the data after processing it from its raw or original form to the form required in the applied algorithm. It contains a transaction ID, Item Name, and reason to know if there is a relationship between the item and specific reasons for return. Each reason contains a Boolean data type with the "T" symbol, which means this item has returned a product with this reason, and the "F" symbol, which means this item hasn't produced a product with this reason.

*Table 4: Shows pre-process of data between items and returns reason.*

| Return Number | Item Name | Return Reason Type 1 | Return Reason Type 2 | Return Reason Type 3 | Return Reason Type N |
|---|---|---|---|---|---|
| 001 | Item 1 | T | F | F | F |
| 002 | Item 2 | F | F | T | F |
| … | … | … | … | … | … |
| n | Item (n) | T | F | F | F |

Table 5 below shows the shape of the data after processing it from its raw or original form to the form required in the applied algorithm. It contains a transaction ID, Month, and reason to know if there is a relationship between the month and specific reasons for return. Each reason contains a Boolean data type with the "T" symbol, which means this month has returned a product with this reason, and the "F" symbol, which means this month hasn't returned a product with this reason.

*Table 5: Shows pre-process of data between the month and returns reason.*

| Return Number | Month | Return Reason Type 1 | Return Reason Type 2 | Return Reason Type 3 | Return Reason Type N |
|---|---|---|---|---|---|
| 001 | Jan | T | F | F | F |
| 002 | Feb | F | F | T | F |
| … | … | … | … | … | … |
| n | December (n) | T | F | F | F |

Table 6 below shows the shape of the data after processing it from its raw or original form to the form required in the applied algorithm. It contains a transaction ID, governorate, and reason to know if there is a relationship between the governorate and specific reasons for return. Each reason contains a Boolean data type with the "T" symbol, which means this governorate has returned a product with this reason, and the "F" symbol, which means this governorate, has not produced a product for this reason.

*Table 6: Shows pre-process of data between governorate and returns reason.*

| Return Number | Governorate | Return Reason Type 1 | Return Reason Type 2 | Return Reason Type 3 | Return Reason Type N |
|---|---|---|---|---|---|
| 001 | Cairo | T | F | F | F |
| 002 | Giza | F | F | T | F |
| … | … | … | … | … | … |
| n | Aswan | T | F | F | F |

Table 7 below shows the shape of the data after processing it from its raw or original form to the form required in the applied algorithm. It contains a transaction ID, day of the week, and reason to know if there is a relationship between the return sales reasons and a specific day in the week. Each reason contains a Boolean data type with the "T" symbol, which means this day of the week has returned a product with this reason, and the "F" symbol, which means this governorate hasn't returned a day of the week with this reason.

*Table 7: Shows pre-process of data between the day of the week and the returns reason.*

| Return Number | Day | Return Reason Type 1 | Return Reason Type 2 | Return Reason Type 3 | Return Reason Type N |
|---|---|---|---|---|---|
| 001 | Sat | T | F | F | F |
| 002 | Sun | F | F | T | F |
| … | … | … | … | … | … |
| n | Friday | T | F | F | F |

## 2.4 Modeling

The IBM SPSS Modeler program was used in this study, which saves us time and effort in the analysis process. Its use is one of the easy programs for researchers, and its results are known for their high accuracy. After the data preparation process, we apply the apriori algorithm since it provides us with a strong relation between categorical data to extract rules.

For the association technique, the Apriori algorithm was used to trace out a relation between (item, customer, plant, day of the week, month, governorate) and specific reasons of return to avoid them in the future EMISAL Company.

The apriori algorithm was used on the IBM SPSS Modeler program, as this algorithm is concerned with extracting the relationships between the data and each other, and by setting a threshold such as the support and confidence coefficient to consider the importance of these relationships or not. Figure 3 below shows the architecture of the purposed data model process using the apriori algorithm under the SPSS Modeler software.

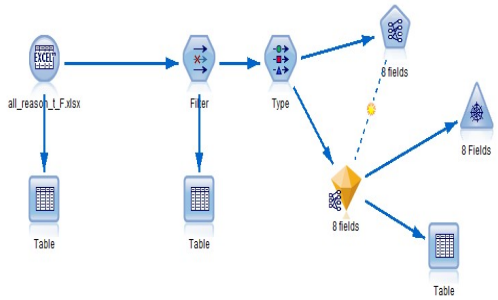


*Figure 3: Shows the proposed model in SPSS Modeler.*

## 2.5 Evaluation

To evaluate and test the result and rules that developed the association apriori model generated by the SPSS Modeler software, the minimum support and confidence values were applied after discussion with the business domain expert to eliminate weak rules, as shown in Table 8 below.

*Table 8: Shows support and confidence Threshold of rules.*

| Model | Support Threshold | Confidence Threshold |
|---|---|---|
| Return and customers | 10 | 20 |
| Return and day | 10 | 20 |
| Return and governorate | 10 | 20 |
| Return and items | 10 | 20 |
| Return and months | 10 | 10 |
| Return and plants | 10 | 10 |

Journal of Theoretical and Applied Information Technology
15th December 2021. Vol.99. No 23
© 2021 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### 2.6 Deployment

At this stage, the results found during the data mining process are used to create a graphical interface like a dashboard that helps the decision-makers in EMISAL to know the reasons for the returns, to use this model in the future by entering new data every period.

## 4. RESULTS

To identify strong rules discovered in the data set, we used different measures called Support and Confidence.

$$\text{Support} = \frac{\text{Frequency}(x,y)}{N} \quad (1)$$

$$\text{Confidence} = \frac{\text{Frequency}(x,y)}{\text{Frequency}(x)} \quad (2)$$

Through this study, the results indicate a set of relationships and correlations between the data and the reasons for its return and there is no doubt that each rule will benefit the company to work with a new strategy to reduce sales returns in the future. The apriori algorithms' results show a strong relation between (day of the week, month, plant, governorate, and customers) as shown in Tables below 9, 10, 11, 12, and 13.

In this study, more than one experiment was done, which resulted in the extraction of different sets of relationships and correlations between the elements. Table 9 below shows the relation between the day of the week and the reason for sales return.

*Table 9: Shows results relation between day of the week and reasons*

| Consequent | Antecedent | Support % | Confidence % |
|---|---|---|---|
| Tuesday | Unmatched weight | 13.70967742 | 35.29411765 |
| Wednesday | Damaged packing materials | 12.09677419 | 33.33333333 |
| Friday | Rain | 13.70967742 | 29.41176471 |

A result in table 9 shows that: -

a. Tuesday has a 35.29 probability for unmatched weight occurring.
b. Wednesday has a 33.33 probability for unmatched weight occurring.
c. Friday has a 29.41 probability for unmatched weight occurring.

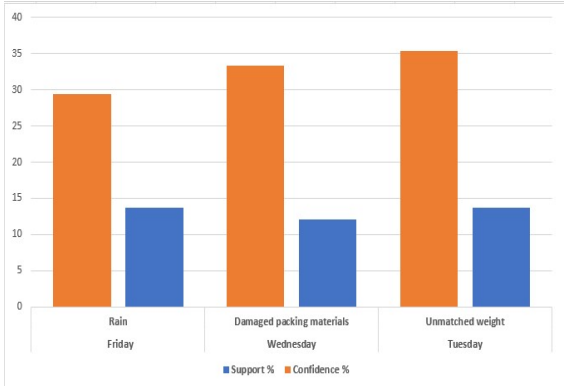


*Figure 4: Shows a bar chart of support & confidence relation between the day of the week and reasons.*

Table 10 below shows the relation between the month of the year and the reason for sales return.

*Table 10: shows results relation between months and reasons*

| Consequent | Antecedent | Support % | Confidence % |
|---|---|---|---|
| The specification does not match | December | 12.09677419 | 80 |
| The specification does not match | November | 15.32258065 | 63.15789474 |
| The specification does not match | October | 10.48387097 | 46.15384615 |
| The specification does not match | March | 10.48387097 | 38.46153846 |

A result in table 10 shows that: -

a. The specification does not match an 80.0 probability for occurring in December.
b. The specification does not match a 63.15 probability for occurring in November.
c. The specification does not match a 46.15 probability for occurring in October.
d. The specification does not match a 38.46 probability for occurring in March.

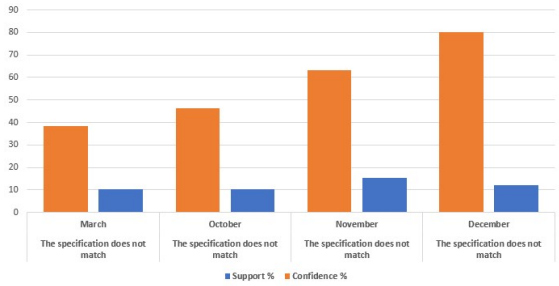


*Figure 5: Shows a bar chart of support & confidence relation between months and reasons.*

Table 11 below shows the relation between the plant of product and the reason for sales return.

*Table 11: Shows results relation between plant id and reasons*

| Consequent | Anteceden t | Support % | Confidenc e % |
|---|---|---|---|
| The specificatio n does not match | Plant 3000 | 22.5806451 6 | 53.5714285 7 |
| The specificatio n does not match | Plant 5000 | 15.3225806 5 | 52.6315789 5 |
| The specificatio n does not match | Plant 2000 | 57.2580645 2 | 52.1126760 6 |

A result in table 11 shows that:-

a. The specification does not match a 53.57 probability for occurring in Plant id number 3000.
b. The specification does not match a 52.63 probability for occurring in Plant id number 5000.
c. The specification does not match a 52.11 probability for occurring in Plant id number 2000.

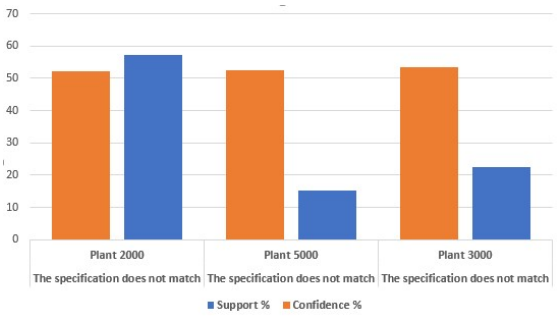


*Figure 6: Shows a bar chart of support & confidence relation between plant id and reasons.*

Table 12 below shows the relation between the governorate and the reason for sales return.

*Table 12: Shows results relation between governorate and reasons*

| Consequent | Anteceden t | Support % | Confidenc e % |
|---|---|---|---|
| 6th of October city | The specificatio n does not match | 50 | 69.3548387 1 |
| 6th of October city | Unmatched weight | 13.7096774 2 | 64.7058823 5 |
| Alexandria | Rain | 13.7096774 2 | 58.8235294 1 |

A result in table 12 shows that:-

a. 6th of October city has a 69.35 probability for occurring with the specification does not match.
b. 6th of October city has a 64.70 probability for occurring with unmatched weight.
c. Alexandria city has a 58.82 probability of occurring with the rain.

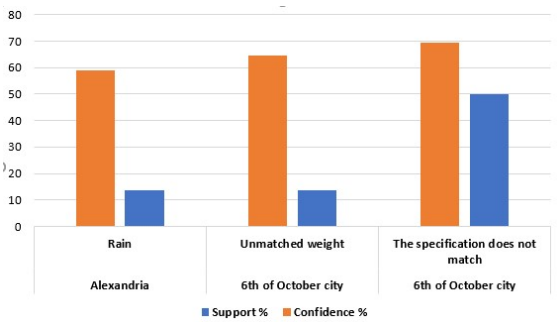


*Figure 7: Shows a bar chart of support & confidence relation between governorate and reasons.*

Table 13 below shows the relation between the customer and the reason for sales return.

*Table 13: shows results relation between customers and reasons*

| Consequent | Antecedent | Support % | Confidence % |
|---|---|---|---|
| The specification does not match | Nestle Egypt | 13.70967742 | 70.58823529 |
| The specification does not match | Crown of Kings Food Industries | 20.16129032 | 48 |
| Rain | Unilever Mashreq Food Products | 16.12903226 | 45 |
| Unmatched weight | Crown of Kings Food Industries | 20.16129032 | 28 |

Results in table 13 show that: -

a. The specification does not match has a 70.58 probability for occurring with customer Nestle Egypt.
b. The specification does not match has a 48.0 probability for occurring with customer Crown of Kings Food Industries.
c. The specification does not match has a 45.0 probability for occurring with customer Unilever Mashreq Food Products.
d. The specification does not match has a 28.0 probability for occurring with customer Crown of Kings Food Industries.

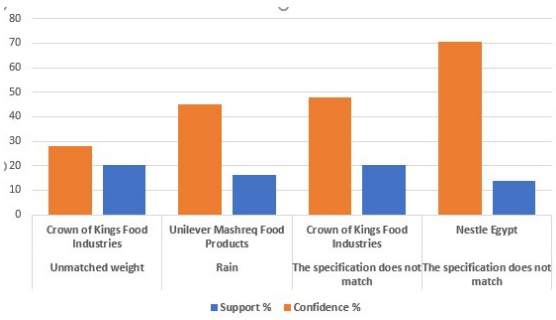


*Figure 8: Shows a bar chart of support & confidence relation between customers and reasons.*

## 5. DISCUSSION

We have selected the most reliable results from each experiment. By reviewing the results, it was found that there are relationships between sales returns between certain days in the week, certain months, a specific factory, a specific product, and a specific customer, and we recommend trying to reduce the occurrence of these cases in the future so that returns are reduced and thus Increase in sales.

Most of the previous studies that worked used the Apriori algorithm were used to extract relationships between items to be used in making offers, the method of this research paper is slightly different as it discusses any relationships between items, months, customers, or factories to know the causes of product returns and work to reduce their occurrence.

## 6. CONCLUSION AND FUTURE WORK

In this study, we analyzed the sales returns in the EMISAL company using a data mining apriori algorithm, and the results indicate the existence of strong relationships between the causes of the returns, products, customers, and governorate of certain days of the week and certain months of the year, put in our consideration these relationships to decrease their occurrence in the future.

As we know that any research cannot be completely the best thing, so in the future study, we will try to collect data on an indirect relationship with the return of sales, such as the production process using data mining.

## ACKNOWLEDGMENT

I would like to acknowledge MEAPAL Data Science @ CORALIA https://www.d-fi.fr/ for their wonderful collaboration and support and all of the opportunities were given to further research.

EMISAL Company supported this research. We also thank this company for providing the data to us to do this research, and on top of that, we value the Chairman and Managing Director of this company, Dr.Abd-Allatif Alkordy.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

**REFRENCES:**

[1] Zulfikar, Wildan Budiawan & Wahana, Agung & Uriawan, Wisnu & Lukman, Nur. Implementation of association rules with apriori algorithm for increasing the quality of promotion.pp 1-5, 2016.

[2] Yong, Wang & Hengshan, Wang. Research and application of small and medium-sized manufacturing enterprises sales promotion model based on data mining. 8th International Conference on Service Systems and Service Management - Proceedings of ICSSSM'11, pp 1-6, 2011.

[3] Setiabudi, Djoni & Purnama, I & Noertjahyana, Agustinus. Data mining market basket analysis' using hybrid-dimension association rules, case study in Minimarket X, pp 196-199, 2011.

[4] Yunyan, Li & Juan, Chen. Application of Association Rules Mining in Marketing Decision-Making Based on Rough Set,pp 3749-3752, 2010.

[5] Maske, Anisha & Joglekar, Bela. Survey on Frequent Item-Set Mining Approaches in Market Basket Analysis.pp 1-5, 2018.

[6] Insani, Rokhmatul & Laksmiwati, Hira.Data mining for marketing in telecommunication industry.pp 179-183, 2016.

[7] Mathur, Robin. Market basket analysis using artificial neural network,pp 1-6, 2014.

[8] Abbas, Wan & Ahmad, Nor & Zaini, Nurlina. Discovering Purchasing Pattern of Sport Items Using Market Basket Analysis.pp 120-125, 2013.

[9] Nengsih, Warnia. A comparative study on market basket analysis and apriori association technique.pp 461-464, 2015.

[10] Pathak, Pankaj & Srivastava, R.R. & Sharma, Ojasvi. Assessment of legislation and practices for the sustainable management of waste electrical and electronic equipment in India. Renewable and Sustainable Energy Reviews.pp 220–232, 2017.

[11] Swamidason, Iwin Thanakumar Joseph. (2019). SURVEY OF DATA MINING ALGORITHM'S FOR INTELLIGENT COMPUTING SYSTEM. Journal of Trends in Computer Science and Smart Technology.pp 14-23. 10.36548/jtcsst.2019.

[12] Bashar, Abul. INTELLIGENT DEVELOPMENT OF BIG DATA ANALYTICS FOR MANUFACTURING INDUSTRY IN CLOUD COMPUTING. Journal of Ubiquitous Computing and Communication Technologies.pp 13-22, 2019.

[13] Abdalla Yasin Mabrouk, Mohamed Hasan Ibrahim, Mohamed Helmy Khafagy, "Improve Sales In Manufacturing Using Data Mining Techniques.", International Journal of Scientific & Technology Research, "IJSTR", December 2019, pp 3657 - 3661.

[14] Isa, Norulhidayah & Kamaruzzaman, N & Ramlan, Muhammad & Mohamed, Nagham & Puteh, Mazidah. Market Basket Analysis of Customer Buying Patterns at Corm Café. International Journal of Engineering and Technology, pp. 119-123, 2018.

[15] Setiawan, Alexander & Budhi, Gregorius & Setiabudi, Djoni & Djunaidy, Ricky. Data Mining Applications for Sales Information System Using Market Basket Analysis on Stationery Company, pp. 337-340, 2017.

[16] W. B. Zulfikar, A. Wahana, W. Uriawan, and N. Lukman, "Implementation of association rules with apriori algorithm for increasing the quality of promotion," 2016 4th International Conference on Cyber and IT Service Management, Bandung, 2016, pp. 234-238.

[17] Pandian, dr. artificial intelligence application in smart warehousing environment for automated logistics. Journal of Artificial Intelligence and Capsule Networks. pp 63-72, 2019.

[18] Chandra, J & Dewi, K. Implementation of Data Mining Sales of Milk Using Apriori Algorithm Method. IOP Conference Series: Materials Science and Engineering, pp 022077, 2019.

[19] Supriyati, Supriyati & Abdillah, S. Data Mining in Sales Data Grouping. IOP Conference Series: Materials Science and Engineering, p. 012116, 2020.

[20] Lu, Xinzhe & Zhang, Haolan & Huang, Ke. Improving Investment Return Through Analyzing and Mining Sales Data,pp 112-118, 2020.

[21] An improved technique for increasing availability in Big Data replication Kaseb, M.R., Khafagy, M.H., Ali, I.A., Saad, E.M.Future Generation Computer Systems, 2019, 91, pp. 493–505

[22] Hash semi cascade join for joining multi-way map reduce Mohamed, M.H., Khafagy, M.H.

[23] IntelliSys 2015 - Proceedings of 2015 SAI Intelligent Systems Conference, 2015, pp. 355–361, 7361167

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[24] Queue weighting load-balancing technique for database replication in dynamic content web sites Sarhan, E., Ghalwash, A., Khafagy, M. Proceedings of the 9th WSEAS International Conference on Applied Computer Science, ACS '09, 2009, pp. 50–55

[25] JOMR: Multi-join optimizer technique to enhance map-reduce job Shanoda, M.S., Senbel, S.A., Khafagy, M.H. 2014 9th International Conference on Informatics and Systems, INFOS 2014, 2015, pp. PDC80–PDC87, 7036682