Application of Data Mining in CRM Market Basket Analysis in Online Retail Dataset

Şeniz Akbulut*

Departmant of Computer Science Engineering , Dokuz Eylül University , İzmir , TURKEY
seniz.akbulut@ceng.deu.edu.tr

January 24, 2022

Abstract

CRM is an iterative process that turns customer data into customer loyalty. In the analysis of this data, data mining techniques are essentially used. Association rules are one of the most frequently used methods which are the special application areas of the data mining. Association rules are the rules that include which items commonly occur together in the same transactions.The Apriori algorithm is the most popular association rule algorithm which discovers all frequent itemsets in large database of transactions. This algorithm uses iterative approach to count the frequent itemsets. The aim of this study is to propose a base for the customer relationship management activities by using data mining tools and applications for a firm in retail sector.

I. Introduction

Customer relationship management is the combination of practices, strategies and technologies that companies use to manage and analyze customer interactions and data throughout the customer lifecycle.It tries to understand the customer profile of the company and to communicate according to these profiles. Moreover CRM targets to gain new customers other than existing customers.CRM is defined as customer relationship management and is a software system that helps business owners nurture their relationships with their clientele. A CRM also assists with organization, efficiency, time management, and impressing clients every step of the way.CRM software has been around since the mid-1990s, but has come into its own over the last decade. CRM platforms are powerful systems that connect all the data from your sales leads and customers all in one place. A CRM records and analyzes all calls, emails and meetings, helping improve customer service, drive sales, and increase revenue.

In today's world, hard conditions in the market lead the companies to find new ways to compete better. With the intensive global competition and rapidly changing technological environments, meeting customers various needs and maximizing the value of profitable customers are becoming the only viable option for many contemporary companies. Customer Relationship Management (CRM) provides organizations with the platform to obtain a competitive advantage by embracing customer needs and building value driven long-term relationships. CRM is an iterative process that turns customer data into customer loyalty. Analyzing the customer database and convert the data into information that will help company develop programs for building customer loyalty. In the analysis of this data, data mining techniques are essentially used. Association rules are one of the most frequently used methods which are the special application areas of the

*I would like to thank Prof.Dr.Recep Alp KUT and his Assistant for their contribution.

data mining. Association rules are the rules that include which items commonly occur together in the same transactions. The Apriori algorithm is the most popular association rule algorithm which discovers all frequent itemsets in large database of transactions. This algorithm uses iterative approach to count the frequent itemsets. Using this algorithm, candidate patterns which receive sufficient support from the database and the algorithm uses aprior gen actions join and prune to find all frequent itemsets. The aim of this study is to propose a base for the customer relationship management activities by using data mining tools and applications for a firm in retail sector. Customer master data and sales transactions of customers are converted to meaningful information that can be used for customer relationship management activities. In this concern, a market basket analysis is performed, and an application was conducted to find association rules from market datasets by using apriori algorithm.

II. LITERATURE REVIEW

i. Data Mining

Data Mining is often defined as "write many of the reports and queries". But in fact the data mining activities do not report creation and query altogether. Data mining is done with a special tool, which executes data operations that have been defined based on the model of analysis. Data mining is the extraction of information or patterns that are important or interesting from the data residing on large data base which had been unknown but potentially useful information [1] Data Mining analyze the data to find information hidden in large amounts of stored data. Data mining is a process that is different from the usual statistical analysis. Function-common functions that are applied to data mining are: [2] - Association, is the process of finding associative rule between a combination of items. - Sequence, the process of finding associative rule between a combination of items at a time and apply more than one period. - Clustering, is the process of grouping a number of data / objects, into groups so that each group of data containing similar data. - Classification, the discovery process model or function or distinguish concepts or classes of data, with the aim to be able to estimate the class of an object that the label is not known. - Regression, is the process of estimating the value predicted by the patterns in the data set. - Solution, is a process of discovery of the root problem and solving of the problems faced by business or at least as information in decision making.

ii. Market Basket Analysis

Market basket analysis is an association in data mining to find attributes that appear in one time [4]. This process can determine customer buying patterns by finding relationships between different items in a sales transaction. The results that have been obtained can be used by retail companies such as sales company / supermarket develop a marketing strategy to see the items that may be purchased simultaneously by consumers. Analysis of the association or the association rule is a data mining techniques to discover the rules of associative between a combination of items [3]. Examples of associations that possibility, namely how likely buyers buy bread with jam simultaneously. Association rules can be determined by two parameters, support and confidence. Support that is a percentage of the value of supporting combinations of items in the database, while confidence is certainty value the strong relationship between items in the rules of association.

iii. Association Rules

Association rules are usually expressed in the form: bread, butter $\rightarrow$ milk (support = 40% confidence = 75%). This rule means 75% of transactions includes bread, butter also

contains milk. While 40% of all transactions containing the third item. Association analysis is defined as a process to find all association rules that meet the minimum requirements to support (minimum support) and the minimum requirements for confidence (minimum confidence) [3]. If support> = minimum support and confidence> = minimum confidence, then the rule can be expressed as interesting rule [2].

iv. Apriori Algorithm

Apriori algorithms including the type of association rules in data mining. In addition to Apriori, which is included in this group are the method and the Generalized Rule Induction Based Hash Algorithm. Association analysis is data mining techniques to discover an association rule between a combination of items. Examples of association rules from the analysis of the purchase in a supermarket is able to know how big the possibility someone to buy bread together with milk. Based on this knowledge, the owner can self-regulate the placement of the goods or designing a marketing campaign by using a combination of discount coupons for certain items. [3] Based on the parameters of support (supporting values) and confidence (value certainty), the association basic methodology can be divided into two phases, namely: Analysis of the highest frequency pattern and establishment of association rules.Algorithm pseudocode can be seen on Fig 1.

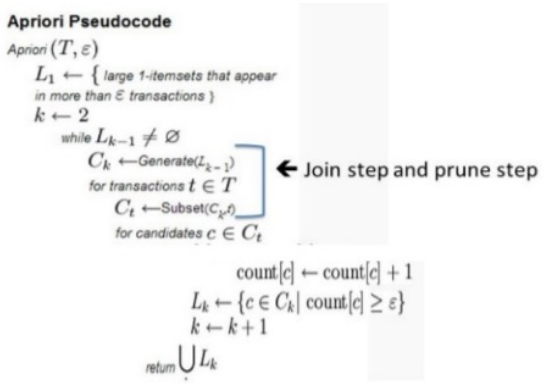

Figure 1: *Pseudo Code of Apriori Algorithm.*

III. Method

There are different algorithms to be used for Association Rules Mining. One of them is the Apriori algorithm. In this project, product association analysis will be handled with "Apriori Algorithm" and the the purchasing tendency of the customer will be revealed who is in the sales process, using the sales data of an e-commerce company.The project was carried out using the UCI Online Retail II dataset.The details of the dataset and solution architecture will be discussed.

i. Business Understanding

Businesses track customer purchasing movements during the process of defining customer habits. In this point, they utilize from association methods. Based upon the purchasing movements, answers are sought for the questions such as which products are sold together with which products, and what similarities exist between customer groups that buy these products.

In sales / marketing campaigns organized by the firm, it is very important to determine which group of customers is going to be directed and also which group of products is coming to the forefront in the process of developing a solution proposal on how to combine the best-selling product with the least-sold product. From this point forth, the basic point to be addressed in the application will be to discover which dual, triple, quadruple combinations from the products are purchased together in the company's product range. In this way, these associations can be taken into consideration in marketing campaigns, new sale targets can be created on a product basis by bringing together the products that are sold more and the products that are less sold, so that these product combinations may be included in the catalogs to be submitted to the customers.

Likewise, the customers whose purchasing behaviors were examined can be divided into groups. Thus, different sales policies can be offered to the customers in different groups.

By identifying the risky customers and also the customers who meet most of the revenue from payment statements, it is possible to make progress in the light of this information on customer relations and sales policies to be developed.

ii. Data Understanding

Dataset is taken from UCI Machine learning repository (https://archive.ics.uci.edu/ml/machine-learningdatabases/00502/).Online Retail II data set contains all the transactions occurring for a UK-based and registered, non-store online retail between 01/12/2009 and 09/12/2011.The company mainly sells unique all-occasion gift-ware. Many customers of the company are wholesalers.

Algorithms such as Apriori, Fp growth, Random forest regressor, linear regression, Ridge regression, Lasso regression can easily use this data set.It allows us to predict the sales of items or predict previously purchased items and make inferences such as the user's probability of purchasing the same items in their next order.There is no missing data in the Data Set.The data set consists of 525461 records and 8 columns.These fields are as listed below.

1. InvoiceNo: A 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'C', it indicates a cancellation (Nominal)
2. StockCode: A 5-digit integral number uniquely assigned to each distinct product (Nominal)
3. Description: Product (item) name. (Nominal)
4. Quantity: The quantities of each product (item) per transaction (Numeric)
5. InvoiceDate: The day and time when each transaction was generated (Numeric)
6. UnitPrice: Product price per unit in sterling (Numeric)
7. CustomerID: A 5-digit integral number uniquely assigned to each customer (Nominal)
8. Country: Name of the country where each customer resides (Nominal)

IV. Methodology

As mentioned in the previous section, the project was developed using the UCI Online Retail II dataset, which is an Opensource dataset. The project was coded using the python language. The dataset description snapshot is given below:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	1/12/2010 8:26	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	1/12/2010 8:26	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	1/12/2010 8:26	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	1/12/2010 8:26	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	1/12/2010 8:26	3.39	17850.0	United Kingdom

Figure 2: *Snapshot of Dataset Description*

i. Data Preperation

We are checking the data to see if there are any null values first and then planning on dividing the dataset to obtain required columns and use them to draw required objectives. Real world data contains null values and a lot of noise. It has to be preprocessed before using it for further analysis. As a result of the analysis, we confirmed that there is no missing value.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 532619 entries, 0 to 532618
Data columns (total 8 columns):
InvoiceNo      532619 non-null object
StockCode      532619 non-null object
Description    531165 non-null object
Quantity       532619 non-null int64
InvoiceDate    532619 non-null object
UnitPrice      532619 non-null float64
CustomerID     397924 non-null float64
Country        532619 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 32.5+ MB
```

Figure 3: *Information about Data*

In addition, in order to work effectively on the data set, we need to know well what type of data the fields in the data set hold. For this reason, the fields and data types are displayed below.

```
Data columns (total 8 columns):
InvoiceNo      532619 non-null object
StockCode      532619 non-null object
Description    531165 non-null object
Quantity       532619 non-null int64
InvoiceDate    532619 non-null object
UnitPrice      532619 non-null float64
CustomerID     397924 non-null float64
Country        532619 non-null object
dtypes: float64(2), int64(1), object(5)
memory usage: 32.5+ MB
```

Figure 4: *Attributes and Data Types*

ii. Descriptive Statistics

Descriptive statistics are used to describe or summarize the characteristics of a sample or data set, such as a variable's mean, standard deviation, or frequency. Inferential statistics can help us understand the collective properties of the elements of a data sample. Knowing the sample mean, variance, and distribution of a variable can help us understand the world around us.For this reason, we looked at descriptive statistics to learn more about the dataset.The conclusions obtained are as follows.

	Quantity	UnitPrice	CustomerID
count	532619.000000	532619.000000	397924.000000
mean	10.240010	3.847635	15294.315171
std	159.573967	41.758101	1713.169877
min	-9600.000000	-11062.060000	12346.000000
25%	1.000000	1.250000	13969.000000
50%	3.000000	2.080000	15159.000000
75%	10.000000	4.130000	16795.000000
max	80995.000000	13541.330000	18287.000000

Figure 5: *Snapshot of Descriptive Statistics*

iii. Visualization

Unit price distribution is visualized.The sales values of different countries were checked.According to the sales figures, 20 countries with the least sales were visualized.How many quantity of products have been sold online situtaion checked from each country.Visualized with bar plot.Quality of Products sold in all countries except the UK visualized.According to the sales figures, the 20 countries with the least sales were compared.The most common words in the product description list were visualized with the word cloud method.Distribution of Sales in entire globe visualized.Sales of All Countries except UK are listed.Then the obtained values were visualized again.According to the sales figures, the last 20 countries were visualized.

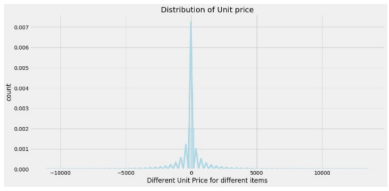

Figure 6: *Visualization of Unit Price Distribution*

```
United Kingdom     487620
Germany              9042
France               8408
EIRE                 7894
Spain                2485
Netherlands          2363
Belgium              2031
Switzerland          1967
Portugal             1501
Australia            1185
Norway               1072
Italy                 758
Channel Islands       748
Finland               685
Cyprus                614
Sweden                451
Unspecified           446
Austria               398
Denmark               380
Poland                330
Name: Country, dtype: int64
```

Figure 7: *Data Counts by Country*

Figure 8: *Buttom 20 Countries*

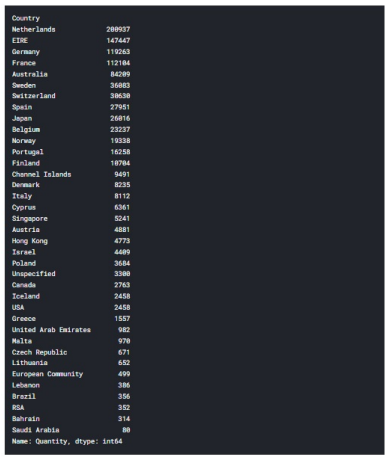

Figure 9: *Quality of Products sold in all the countries except UK*

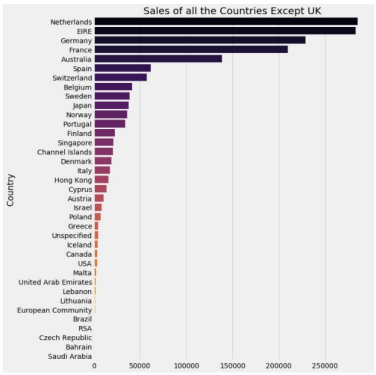

Figure 10: *Quality of Products sold in all the countries except UK*

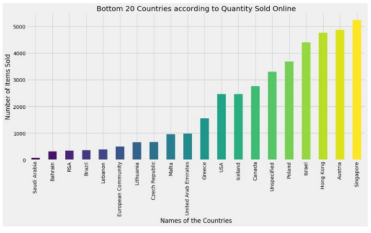

Figure 11: *Having a look at the bottom 20 Countries in terms of Quantities according to the countries*

Figure 12: *Most Occuring word in the Description list*

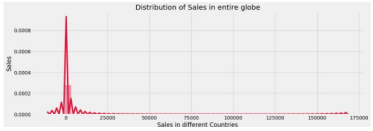

Figure 13: *Sales in different Countries*

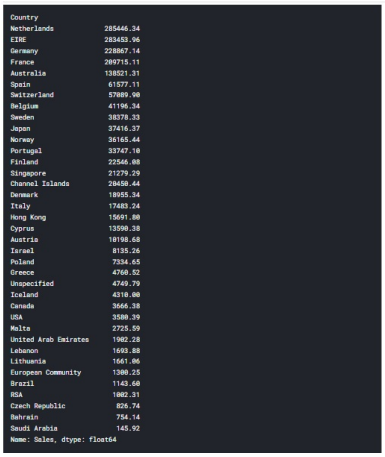

Figure 14: *Each country's Sales*

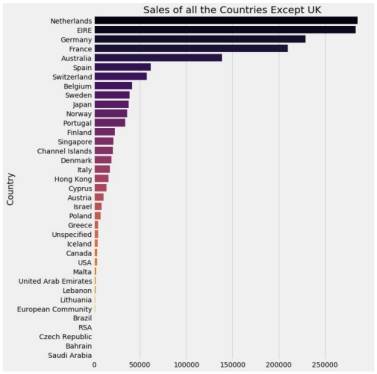

Figure 15: *Sales of all the Countries Except UK*

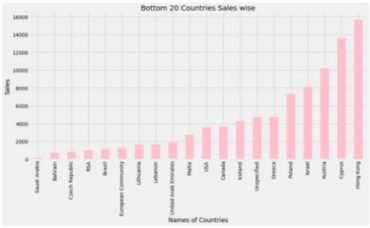

Figure 16: *Looking at the bottom 20 countries sales wise.*

iv. Time Series Analysis

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time. What sets time series data apart from other data is that the analysis can show how variables change over time. In other words, time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. It provides an additional source of information and a set order of dependencies between the data. Time series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting—predicting future data based on historical data.

So ,sales and invoicing dates were evaluated and visualized by time series analysis. The

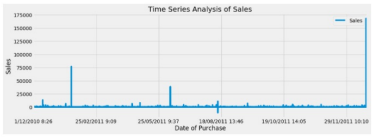

Figure 17: *Sales vs Invoicedate (Time series Analysis)*

number of Unique IDs was calculated.How many unique countries present in this dataset checked.Defined a function to plot time-series plot for any country.Sales and time graphs of all countries were visualized through time series analysis.

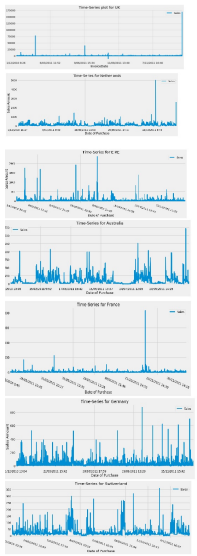

Figure 18: *Time Series Analysis for Some Countries*

v. Data Grouping

The product stock codes were grouped in the data and the number of products was checked. The data set was sorted according to the sales figures. Then it was sorted again by looking at the product prices. The lines that did not have any invoice numbers, that is, the sales were not

completed, were destroyed.The data splitted according to the region of transaction.Dataset encoded.Data baskets were created.

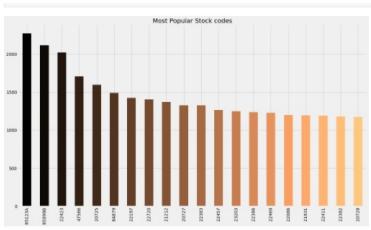

Figure 19: *Looking the stockcode for the datset)*

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Sales
294522	A563186	B	Adjust bad debt	1	12/08/2011 14:51	-11062.06	NaN	United Kingdom	-11062.06
294523	A563187	B	Adjust bad debt	1	12/08/2011 14:52	-11062.06	NaN	United Kingdom	-11062.06
40798	539656	21116	OWL DOORSTOP	1	22/12/2010 14:41	0.00	NaN	United Kingdom	0.00
325066	565938	23066	NaN	-13	8/09/2011 10:54	0.00	NaN	United Kingdom	-0.00
140178	549648	21472	NaN	-140	1/04/2011 13:15	0.00	NaN	United Kingdom	-0.00
...	...	...	...	...	...	...	...	...	...
294521	A563185	B	Adjust bad debt	1	12/08/2011 14:50	11062.06	NaN	United Kingdom	11062.06
14839	537632	AMAZONFEE	AMAZON FEE	1	7/12/2010 15:08	13541.33	NaN	United Kingdom	13541.33
218664	556444	22502	PICNIC BASKET WICKER 60 PIECES	60	10/06/2011 15:28	649.50	15098.0	United Kingdom	38970.00
60580	541431	23166	MEDIUM CERAMIC TOP STORAGE JAR	74215	18/01/2011 10:01	1.04	12346.0	United Kingdom	77183.60
531138	581483	23843	PAPER CRAFT , LITTLE BIRDIE	80995	9/12/2011 9:15	2.08	16446.0	United Kingdom	168469.60

Figure 20: *Sorting the dataset by sales amount)*

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Sales
14839	537632	AMAZONFEE	AMAZON FEE	1	7/12/2010 15:08	13541.33	NaN	United Kingdom	13541.33
294521	A563185	B	Adjust bad debt	1	12/08/2011 14:50	11062.06	NaN	United Kingdom	11062.06
170160	551697	POST	POSTAGE	1	3/05/2011 13:46	8142.75	16029.0	United Kingdom	8142.75
292286	562955	DOT	DOTCOM POSTAGE	1	11/08/2011 10:14	4505.17	NaN	United Kingdom	4505.17
263194	560373	M	Manual	1	18/07/2011 12:30	4287.63	NaN	United Kingdom	4287.63
414584	573080	M	Manual	1	27/10/2011 14:20	4161.06	12536.0	France	4161.06
414560	573077	M	Manual	1	27/10/2011 14:13	4161.06	12536.0	France	4161.06
398897	571751	M	Manual	1	19/10/2011 11:18	3949.32	12744.0	Singapore	3949.32
367877	569382	M	Manual	1	3/10/2011 16:44	3155.95	15502.0	United Kingdom	3155.95
341714	567353	M	Manual	1	19/09/2011 16:14	2653.95	NaN	Hong Kong	2653.95
117455	546556	M	Manual	1	15/03/2011 9:50	2583.76	NaN	Hong Kong	2583.76
292031	562946	M	Manual	1	11/08/2011 9:38	2500.00	15581.0	United Kingdom	2500.00
142170	548813	M	Manual	1	4/04/2011 13:03	2382.92	12744.0	Singapore	2382.92
484322	578149	DOT	DOTCOM POSTAGE	1	23/11/2011 11:11	2275.54	NaN	United Kingdom	2275.54
515827	580610	DOT	DOTCOM POSTAGE	1	5/12/2011 11:48	2196.67	NaN	United Kingdom	2196.67
398898	571751	M	Manual	1	19/10/2011 11:18	2118.74	12744.0	Singapore	2118.74
516069	580612	DOT	DOTCOM POSTAGE	1	5/12/2011 11:58	2114.00	NaN	United Kingdom	2114.00
142173	548820	M	Manual	1	4/04/2011 13:04	2053.07	12744.0	Singapore	2053.07
336841	566927	M	Manual	1	15/09/2011 15:20	2033.10	17846.0	United Kingdom	2033.10
493256	578833	DOT	DOTCOM POSTAGE	1	25/11/2011 15:23	2028.25	NaN	United Kingdom	2028.25

Figure 21: *Sorting the datset by unit price)*

vi. Apriori Algorithm Rule Mining

Apriori algorithm refers to the algorithm which is used to calculate the association rules between objects. It means how two or more objects are related to one another. In other words, we can say that the apriori algorithm is an association rule leaning that analyzes that people who bought product A also bought product B. The primary objective of the apriori algorithm is to create the association rule between different objects. The association rule describes how two or more objects are related to one another. Apriori algorithm is also called frequent pattern mining. Generally, you operate the Apriori algorithm on a database that consists of a huge number of transactions.

Apriori algorithm refers to an algorithm that is used in mining frequent products sets and relevant association rules. Generally, the apriori algorithm operates on a database containing a huge number of transactions. For example, the items customers but at a Big Bazar. Apriori algorithm helps the customers to buy their products with ease and increases the sales performance of the particular store.Thre are thre components comprise the apriori algorithm.

1. Support
2. Confidence
3. Lift

Support refers to the default popularity of any product. You find the support as a quotient of the division of the number of transactions comprising that product by the total number of transactions.Confidence refers to the possibility that the customers bought both item A and item B together. So, you need to divide the number of transactions that comprise both item A and item B by the total number of transactions to get the confidence.Consider the above example; lift refers to the increase in the ratio of the sale of item B when you sell item A. So algorithm uses Four Parameters:

min_support: support refers to the popularity of item and can be calculated by finding the number of transactions containing a particular item divided by the total number of transactions.

Support(diaper) = (Transactions containing (diaper))/(Total Transactions) Support(diaper) = 150 / 1000 = 15 %

min_confidence: Confidence refers to the likelihood that an item B is also bought if item

A is bought. It can be calculated by finding the number of transactions where A and B are bought together, divided by the total number of transactions where A is bought.

Mathematically, it can be represented as: Confidence(A → B) = (Transactions containing both (A and B))/(Transactions containing A)

The confidence of likelihood of purchasing a diaper if a customer purchase milk. Confidence(milk → diaper) = (Transactions containing both (milk and diaper))/(Transactions containing milk) Confidence(milk → daiper) =30 / 120 = 25 %

Confidence is similar to Naive Based Algorithm.

min_lift: Lift refers to the increase in the ratio of the sale of B when A is sold. Lift(A –> B) can be calculated by dividing Confidence(A -> B) divided by Support(B). Mathematically it can be represented as: Lift(A→B) = (Confidence (A→B))/(Support (B))

Lift(milk → diaper) = (Confidence (milk → diaper))/(Support (diaper)) Lift(milk → diaper) = 25 / 15 = 1.66

So by Lift theory, there is 1.66 times more chance of buying milk and diaper together then just buying diaper alone.

min_length: How many Items do we want to associate in our rules.

Rules for Apriori Algorithm

1. Set a minimum value for support and confidence. This means that we are only. interested in finding rules for the items that have certain default existence (e.g. support) and have a minimum value for co-occurrence with other items (e.g. confidence).
2. Extract all the subsets having a higher value of support than a minimum threshold.
3. Select all the rules from the subsets with confidence value higher than the minimum threshold.
4. Order the rules by descending order of Lift.

```
In [32]:
# Building the model
frq_items = apriori(basket_France, min_support = 0.05, use_colnames = True)

# Collecting the inferred rules in a dataframe
rules = association_rules(frq_items, metric ="lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])
rules.head()
```

Out[32]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
45	(JUMBO BAG WOODLAND ANIMALS)	(POSTAGE)	0.076531	0.765306	0.076531	1.000	1.306667	0.017961	inf
258	(PLASTERS IN TIN CIRCUS PARADE, RED TOADSTOOL ...	(POSTAGE)	0.051020	0.765306	0.051020	1.000	1.306667	0.011974	inf
270	(RED TOADSTOOL LED NIGHT LIGHT, PLASTERS IN TI...	(POSTAGE)	0.053571	0.765306	0.053571	1.000	1.306667	0.012573	inf
302	(SET20 RED RETROSPOT PAPER NAPKINS, SET6 RED ...	(SET6 RED SPOTTY PAPER PLATES)	0.102041	0.127551	0.096490	0.975	7.644000	0.084474	34.897959
300	(SET6 RED SPOTTY PAPER PLATES, SET20 RED RET...	(SET6 RED SPOTTY PAPER CUPS)	0.102041	0.137755	0.096490	0.975	7.077778	0.085433	34.489796

Figure 22: *Association Rules)*

V. CONCLUSION

As a result of the study, information about the data set was obtained by applying data mining techniques on the data set, the data was processed and analyzed and a rule inference was made by applying the Apriori algorithm. These inferences were analyzed with the aim of using CRM. The Apriori algorithm is the most known and efficient algorithm for market basket analysis.

References

[1] Rifa, Noverino (2004). Bussiness Intelligence, ITB

[2] Han, Jiawei Kember, Michelin, 2001, Data Mining Concepts Techniques, Simon Fraser University Academic Press, US

[3] Kusrini, Emha T.L., Algoritma Data Mining, Yogyakarta: Penerbit Andi.

[4] giudici, P. and Figini, S. (2009) Market Basket Analysis, in Applied Data Mining for Business and Industry, Second Edition, John Wiley Sons, Ltd, Chichester, UK. doi: 10.1002/9780470745830.ch7.

[5] Setiawan, Alexander Budhi, Gregorius Setiabudi, Djoni Djunaidy, Ricky. (2017). Data Mining Applications for Sales Information System Using Market Basket Analysis on Stationery Company. 337-340. 10.1109/ICSIIT.2017.39.

[6] Durdu, M. (2012). Application of Data Mining in Customer Relationship Management Market Basket Analysis in a Retailer Store. Master Thesis, Izmir: DokuzEylülUniversity, Institute of Natural Sciences.