MEF UNIVERSITY

GİTTİGİDİYOR BASKET ANALYSIS

Capstone Project

Kerem YILMAZ

İSTANBUL, 2017

MEF UNIVERSITY

GİTTİGİDİYOR BASKET ANALYSIS

Capstone Project

Kerem YILMAZ

Advisor: Dr. Tuna ÇAKAR

İSTANBUL, 2017

ACKNOWLEDGEMENTS

I want to thank Dr. Tuna Çakar for helping me all the time with ideas for my thesis and for his contribution to my capstone project. I want to thank Prof. Dr. Özgür Özlük for him contribution to my graduate education. I want to thank Sanem Balkan of Gittigidiyor.com for supplying me the data I used in my capstone project. I want to thank my family who always supported me and never left me alone during the preparation of this project.

EXECUTIVE SUMMARY

GİTTİGİDİYOR BASKET ANALYSIS

Kerem YILMAZ

Advisor: Dr. Tuna ÇAKAR

SEPTEMBER, 2017, 33 pages

Companies have produced masses of data from their customers for decades. The companies are still saving significant amount of data in days not in months or years with the e-commerce applications growing rapidly. However, the companies poor in information extracted from that data. Big Data is the solution of that issue and the companies are relying on it for making strategic decisions.

Market Basket Analysis is a very useful technique for extracting information to data and implementing marketing strategies customized to the individual level. Especially, it is used finding out co-occurring items in consumer shopping basket. Such information can be used as basis for decisions about marketing activity such as promotional support and cross-sale campaigns. Main reason of this, MBA divides the purchases from the store into baskets to determine what items are commonly purchased together. MBA ensure lots of benefits that improves efficient of business model and marketing strategy. One of important benefit is the improvement in revenue based on probability that a product will be demanded at a given price that consumers are willing to pay. One of other important benefit includes data analysis of consumer shopping habits that allow decision-makers within the company to create optimal store layouts and product placement that boost both revenue and customer satisfaction.

Market Basket Analysis finds interesting patterns from databases such as association or correlation relationships among a large set of data items. Association rules are derived from the frequent item sets using support and confidence as threshold levels. The support of an item set is defined as measures the number or proportion of transactions that contain all the items in the rule. Confidence is defined as the measure of certainty or reliability associated with each discovered pattern. MBA is using one of Data Mining algorithms that Apriori Algorithm. Apriori Algorithm is a level-wise, breadth-first algorithm which counts transactions and uses prior knowledge of frequent item set properties.

The main aim of this project we have done with the Gittigidiyor company is to find high-frequency associations in the sale of mobile phones, headphones or other accessories sold in the mobile-phone category and to use this information in marketing. First, we will try to create a model that can find the rules of the relationship using the data provided by the company. The success of a model can be measured by the consumer responding to the product offer offered to him. If the algorithm produces output, the outputs can be presented to the consumer and the performance can be measured. Successful outcomes can also reach other users who have the same behavior and increase company sales rates. In this direction, new campaigns and sales strategies that can increase the company profit rate can be developed.

Key Words: Association rules, market basket analysis, apriori algorithm

ÖZET

GİTTİGİDİYOR SEPET ANALİZİ

Kerem YILMAZ

Tez Danışmanı: Dr. Tuna ÇAKAR

EYLÜL, 2017, 33 sayfa

Birçok şirket yıllar boyunca müşterilerinden bazı verileri topladılar ve biriktirdiler. Bu şirketlere e-ticaret şirketleri de eklenince, veri artık çok daha hızlı bir şekilde birikmeye başladı. Bununla birlikte, şirketler veriyi kullanmakta hayli geri kalmış ve bilgisizdi. Bugün, şirketler stratejik karar alırken bu veri yığınını kullanmak için çalışmalar yapmakta ve bu çalışmaların etkilerine inanmaktadır.

Pazar Sepeti Analizi, verilerden bilgi ayıklamak ve bireysel seviyeye uyarlanmış pazarlama stratejileri uygulamak için çok kullanışlı bir tekniktir. Özellikle, tüketici alışveriş sepetindeki eş zamanlı öğelerin bulunmasında kullanılır. Bu tür bilgiler, promosyon desteği ve çapraz satış kampanyaları gibi pazarlama faaliyetleri ile ilgili kararlar için temel olarak kullanılabilir. Pazar Sepet Analizi'nin ana amacı, alışveriş sepetinde bulunan hangi ürünlerin birlikte satın alındığını keşfedebilmektir. Sepet analizi, iş modeli ve pazarlama stratejisinin etkinliğini artıran birçok avantaj sağlar. Bu avantajlarına örnek olarak, tüketicilerin ödemeyi kabul ettiği belirli bir fiyatla bir ürünün talep edileceğine bağlı olarak gelirdeki iyileşmedir. Diğer önemli faydalardan biri de, karar alıcıların şirket içinde en iyi mağaza düzenleri ve ürün yerleşimi yaratarak hem geliri hem de müşteri memnuniyetini artırabilmesini sağlayan tüketici davranışlarını ölçümleyebilmesidir.

Pazar Sepeti Analizi, veri setinde yer alan ürünler arasındaki korelasyona bakarak bazı kalıplaşmış ürün alımlarını veya ilişki kurallarını keşfeder. İlişki kuralları, destek ve güven olarak adlandırılan değişkenler belirlenerek türetilir. Bir öğe kümesinin desteği, kuraldaki tüm öğeleri içeren işlemlerin sayısını veya oranını ölçmek olarak tanımlanır. Güven ise, keşfedilen her modelle ilgili kesinlik veya güvenilirlik ölçüsü olarak tanımlanır. Sepet analizinin temelinde istatistiksel analiz olmasından dolayı bu güne kadar bir çok farklı algoritma geliştirilmiştir. Bu aigoritmalar içinde en gelişmiş olanı ise Apriori Algoritması olarak bilinmektedir. Apriori Algoritması, breadth-first algoritmasını esas alarak istatistiksel birliktelikleri sayarak sıkça birlikte alınan ürünleri bulmakta çok becerikli bir algoritmadır.

Gittigidiyor şirketi ile birlikte yaptığımız bu projenin temel amacı, cep telefonu kategorisinde satılmış olan cep telefonu, kulaklık veya diğer aksesuarların satışındaki yüksek sıklıktaki birliktelikleri bulabilmek ve bu bilgilerin pazarlama alanında kullanılmasını sağlamaktır. Öncelikle, şirket tarafından sağlanan veriler kullanarak ilişki kurallarını bulabilecek bir model yaratmaya çalışacağız. Bir modelin başarısı, tüketicinin ona sunulan ürün teklifine cevap vermesi ile ölçülebilir. Algoritmanın çıktılar üretmesi durumunda, çıktılar tüketiciye sunulabilir ve başarısı ölçülebilir. Başarılı olan çıktılar sayesinde aynı davranışlara sahip olan diğer kullanıcılara da ulaşılır ve şirket satış oranları artabilir. Bu doğrultuda şirket kar oranını artırabilecek yeni kampanya ve satış stratejileri geliştirilebilir.

Anahtar Kelimeler: İlişki kuralı, alış-veriş sepet analizi, apriori algoritması

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Overview

Data Mining is becoming more important for lots of sector and companies worlwide. Because, it can find patterns, correlations, anomalies in the databases which can help us to make accurate future decision. Data Mining contains of various statistical analyses that reveal unknown aspect of the data. Data Mining encompasses a huge variety of statistical and computational techniques such as; Market Basket Analysis, Clustering, Classification and Regression Analyses.

The collection and study of retail transaction data, known as Market Basket Analysis(MBA), has become increasingly prevalent in the past several years. MBA adresses many of the key challenges that companies face today by answering number of vexing questions. Leading companies are learning yo use MBA to make their bussiness more predictable and profitable by identifying product similarities. These patterns in the transaction datas reveal how customers buy, which is extremely variable across the retail spectrum, even on an e-commerce or a typical grocery store.

1.2. Research Problem Description

In recent years, making analyze shopping basket has become quite appealing to the companies. Today's technology made it possible for them to produce information on their customers and what they buy. Software and applications allow accessing to the shopping transaction data. In retail business analyzing such information is highly useful for understanding buying behavior. Mining purchasing patterns allows companies to adjust promotions, store settings and serve customers better. Transactional data is used for mining valuable information on co-purchases and adjusting promotion and advertising accordingly.

1.3 Literature Review

Data mining has taken an important part of marketing literature for the last several decades. Market basket analysis is one of the important areas in the field of data mining and is the best example for mining association rules. Association Rule Mining identifies the association or relationship between a large set of data items and forms the base for Market Basket Analysis. In Market Basket Analysis, buying habits of customers are analyzed to find associations between the different items that customers place in their shopping baskets. Various algorithms for Association Rule Mining (ARM) and Clustering have been developed by researchers to help users achieve their objectives. Rakesh Agrawal and Usama Fayyad are one of the pioneers in data mining. They account for a number of developed algorithms and procedures. According to Shapiro, rule generating procedures can be divided into procedures that find quantitative rules and procedures that find qualitative rules.

Since the introduction of the Apriori Algorithm, it has been considered the most useful and fast algorithm for finding frequent item sets. Many improvements have been made on the Apriori Algorithm in order to increase its efficiency and effectiveness. (M.J.Zaki, M.Ogihara, S. Parthasarathy, 1996). There are few algorithms developed that are not based on the Apriori, but they still address the issue of speed of Apriori. The following papers (Eu-Hong Sam Han, George Karypis, Vipin Kumar, 1999), (Jong Soo Park, Ming-Syan Chen, Philip S. Yu) propose new algorithms which are not based on the Apriori, but all of them are being compared to Apriori in terms of execution time.

2. ABOUT THE DATA

2.1 Data Description

The given dataset covers the sales records in a huge transactional database. The study is based on data from cell phone category of an e-commerce company that gittigidiyor.com in Istanbul, Turkey. The company sell the products from the e-commerce website to the consumers.

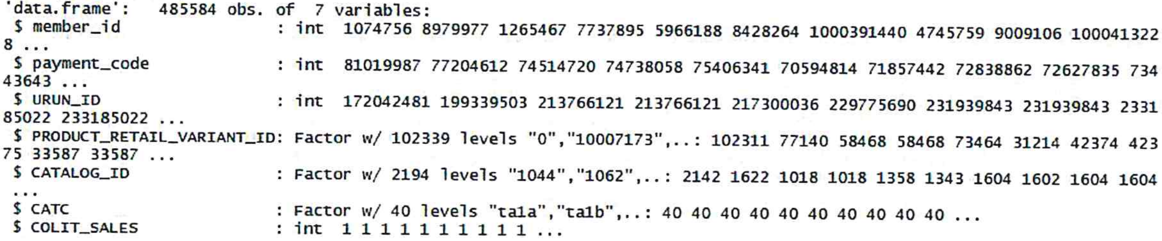
The transaction database consists of the following information:

- Payment code – unique basket identification;
- Member id – unique consumer identification;
- Product id – unique product identification;
- Category id – unique category identification;
- Catalog id – unique catalog identification;
- Product retail variant id
- Colit Sales

In given dataset contains duration of 7 months, cell phones and accessories were aggregated over product categories. The attributes consist of the identifications because of the privacy of the e-commerce company. We are going to analyze the results at the end of the project with the company.

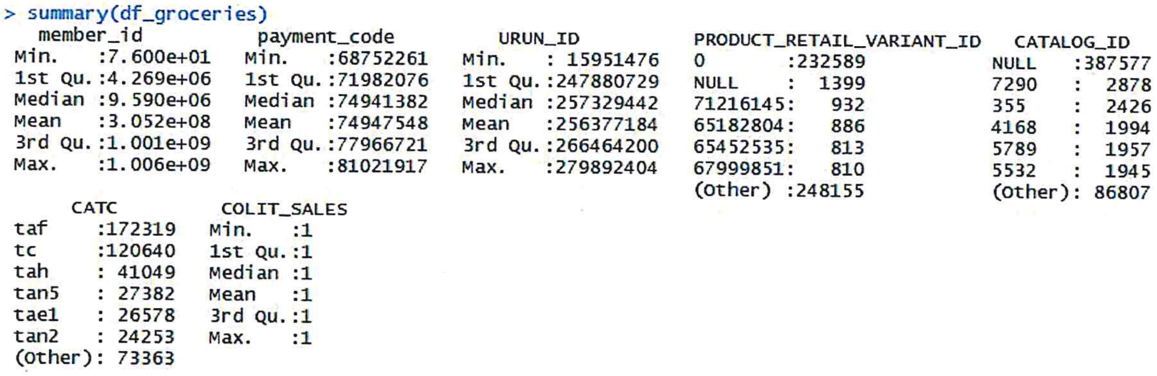
2.2 Data Pre-Processing

The transaction data set can be used to keep events in the form of a file where each record represents the transaction. This section presents a methodology known as market basket transactions, which is useful for discovering interesting relationships hidden in big datasets.



```
> str(df_groceries)
'data.frame':	485584 obs. of  7 variables:
 $ member_id                 : int  1074756 8979977 1265467 7737895 5966188 8428264 1000391440 4745759 9009106 100041322
8 ...
 $ payment_code              : int  81019987 77204612 74514720 74738058 75406341 70594814 71857442 72838862 72627835 734
43643 ...
 $ URUN_ID                   : int  172042481 199339503 213766121 213766121 217300036 229775690 231939843 231939843 2331
85022 233185022 ...
 $ PRODUCT_RETAIL_VARIANT_ID: Factor w/ 102339 levels "0","10007173",..: 102311 77140 58468 58468 73464 31214 42374 423
75 33587 33587 ...
 $ CATALOG_ID                : Factor w/ 2194 levels "1044","1062",..: 2142 1622 1018 1018 1358 1343 1604 1602 1604 1604
 ...
 $ CATC                      : Factor w/ 40 levels "ta1a","ta1b",..: 40 40 40 40 40 40 40 40 40 40 ...
 $ COLIT_SALES               : int  1 1 1 1 1 1 1 1 1 1 ...
```

Figure 1: Data Overview



```
> summary(df_groceries)
   member_id           payment_code            URUN_ID          PRODUCT_RETAIL_VARIANT_ID   CATALOG_ID
 Min.    :7.600e+01   Min.    :68752261   Min.    : 15951476   0         :232589          NULL    :387577
 1st Qu.:4.269e+06   1st Qu.:71982076   1st Qu.:247880729   NULL      :  1399          7290    :  2878
 Median :9.590e+06   Median :74941382   Median :257329442   71216145:    932          355     :  2426
 Mean   :3.052e+08   Mean   :74947548   Mean   :256377184   65182804:    886          4168    :  1994
 3rd Qu.:1.001e+09   3rd Qu.:77966721   3rd Qu.:266464200   65452535:    813          5789    :  1957
 Max.   :1.006e+09   Max.   :81021917   Max.   :279892404   67999851:    810          5532    :  1945
                                                            (Other) :248155          (Other):  86807
      CATC          COLIT_SALES
 taf     :172319   Min.    :1
 tc      :120640   1st Qu.:1
 tah     : 41049   Median :1
 tan5    : 27382   Mean    :1
 tae1    : 26578   3rd Qu.:1
 tan2    : 24253   Max.    :1
 (Other): 73363
```

Figure 2: Data Summary

3. PROJECT DEFINITION

3.1 Problem Statement

Gittigidiyor.com is an ecommerce company that buy lots of different product in different category. We have the data which has 458585 measurement about basket transaction. We will try to identify the best possible combinatory of the products or services which are frequently bought by the customers. The problem is mainly understand that how we increase the basket revenue. If we recommend correct product to correct consumer we can success for the shopping basket revenue. To develop an efficient algorithm to find the desired information resources and their usage pattern and also to develop an algorithm for geographical data sets that reduces communication cost and communication overhead. This project is aimed at designing and implementing a well-structured market basket analysis software tool to solve the problem stated above and compare the result to that of an existing software called R.

3.2 Problem Objectives

It has become increasingly necessary for users to utilize automated tools in find the desired information resources, and to track and analyze their usage patterns. Association rule mining is an active datamining research area. However, most Association rule mining algorithms cater to a centralized environment. Therefore, Rule Mining algorithms have been developed.

3.3 Project Scope

This scope of the study focuses on gittigidiyor.com e-commerce and the scope of this project includes:

- We aim to develop our very own market basket analysis algorithm
- The analysis will be based on Apriori Algorithm.
- We aim to find the consumer who make similar shopping.
- We aim to increase the basket revenue with offering and recommending correct product to correct customer

4. MEDHODOLOGY

4.1 Association Rule Mining

Using data mining techniques on transactional data leads to the generation of association rules and finding correlations between products in the records. The main concept of association rules is to examine all possible rules between items and turn them into 'if-then' statements.

4.1.1 Definition

Let I = { i1,i2,i3,. . . , im} is the set of all items available at the store.

By T = {t1,t2,t3, . . . , tn} we define the set of all transactions in the store.

Each transaction ti = {i2,i4,i9} contains a subset of items from the whole market basket dataset.

An itemset is every collection of zero or more items from the transaction database.

The number of items that occur in a transaction is called a transaction width.

Let's suppose X is a set of items, e.g. X = {Iphone 5s, Samsung j7, Samsung s7}

Transaction tj contains an item set X if X is a subset of tj(X⊑ tj).

An association rule can be expressed in the form of X →Y, where X and Y are two disjoint item sets (do not have any items in common). X is an antecedent and Y is a consequent, in other words, X implies Y.

The main concept of association rules is to examine all possible rules between items and turn them into "if-then" statements. In this case the "if" part is X or the antecedent, while the 'then' part is Y or the consequent.

Antecedent → consequent [support, confidence]

The antecedent and consequent are often called rule body and rule head accordingly. The generated association rule relates the rule body with the rule head. There are several important criteria of an association rule: the frequency of occurrence, the importance of the relation and the reliability of the rule.

Function	Definition
Support	$S(X \rightarrow Y) = P(X,Y)$ and $S(X) = P(X)$
Confidence	$C(X \rightarrow Y) = P(Y \mid X)$
Lift	$L(X \rightarrow Y) = c(X \rightarrow Y) / P(Y) = P(X,Y) / (P(X)P(Y))$

Table 1: Functions of Association Rules

There are two basic parameters of Association Rule Mining (ARM): support and confidence. They both measure the strength of an association rule. Since the database is quite large, there is a risk of generating too many unimportant and obvious rules, which may not be of our interest. In that case a common practice is to define thresholds of support and confidence prior to analysis if we want to generate only useful and interesting rules.

Support of an association rule is the percentage of records that contain X U Y to the total number of records in the database. In other words, the support measures how often a rule is applicable to the given dataset. In this measure of strength, quantity is not taken into account. The support count increases by one for each time the item is encountered in a different transaction T from the database. For example, if a customer buys three tubes of Iphone 5 headphones in a single transaction, the support count number of [Iphone 5 headphones] increases by one.

Support can be derived from the following formula:

$$\text{Support (XY)} = \frac{Support\ count\ of\ XY}{Total\ number\ of\ transactions}$$

Confidence of an association rule is defined as the percentage of the number of transactions that contain XUY to the total number of records that contain X. In other words, confidence is a measure of the strength of association rules and is used to determine how frequently items from item set Y appear in transactions that contain item set X. Let's suppose we have a rule X→Y. Confidence tells us how likely it is to find Y in a transaction that contains X.

Confidence can be derived from the following formula:

$$\text{Confidence (X/Y)} = \frac{Support\ (XY)}{Support\ (X)}$$

Lift measures the importance of a rule. The lift value is represented as the ratio of the confidence and the expected confidence of a rule. The lift can take over values between zero and infinity. In every association rule we have an antecedent and a consequent, also called rule body and rule head accordingly. If the value of the lift is greater than 1 this means that both the rule body and the rule head appear more often together than expected. The occurrence of the rule body positively affects the occurrence of the rule head. The other way around, if the lift value is lower than 1, this means that both the rule body and rule head appear less often together than expected and the occurrence of the rule body negatively affects the occurrence of the rule head. However, if the lift value is near 1, the rule body and rule head appear together as often as expected.

Lift can be derived from the following formula:

$$L(X \rightarrow Y) = c(X \rightarrow Y) / P(Y) = P\ (X,Y) / (P(X)P(Y))$$

The association rules problem can be easily defined as it follows:

Given a threshold S (the minimum support) and a threshold c (the minimum confidence), we are interested to find all rules in the form of X → Y, where X and Y are sets of items, such that:

1. X and Y appear together in at least s% of the transactions.
2. Y occurs in at least c% of the transactions, in which X occurs.

A given association rule is supported in the database, if it meets both the minimum support and minimum confidence criteria. The main purpose of Association rule mining is to find items that satisfy the prerequisite conditions for minimum support and minimum confidence.

4.2 Frequent Item set Generation

This first step in affinity analysis consists of generating all the rules that would be candidates for indicating association between the items. In other words, the idea is to find all possible combinations of single items, pairs of items, triplets of items and so on in the transactional database. However, as already mentioned, as the number of items and therefore possible combinations increases, the level of complexity rises exponentially. A dataset with k items can potentially generate up to 2k -1 frequent item sets. A lattice structure is usually used to visualize all possible combinations of items in frequent item sets.

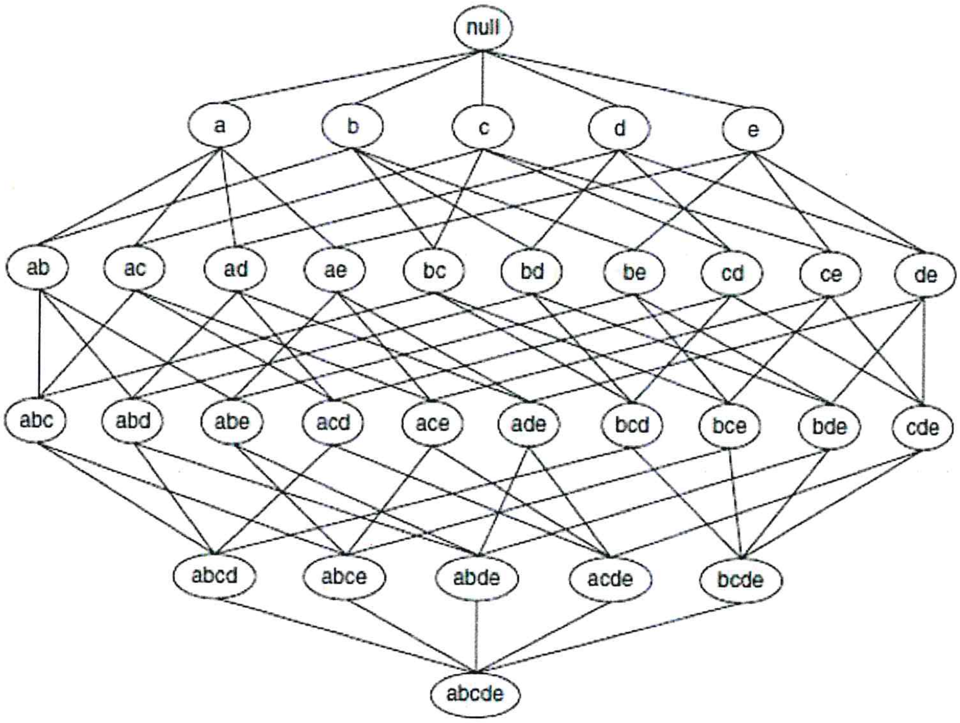


Figure 3: Most Frequent Item Diagram

In order to determine the support count for every candidate item set we need an efficient technique that can find an optimal solution. The classic approach for generating frequent item sets is using the Apriori algorithm. According to the Apriori property: ‘All subsets of a frequent item set must also be frequent'. If it has been verified that an item set X is infrequent, there is no need for further investigating its subsets as they must be infrequent too.

4.3 The Apriori Algorithm

The Apriori is the most commonly used algorithm for frequent item set mining. It starts with identifying the frequent individual items in the transactional database and proceeds with extending them to larger and larger item sets until they appear often enough in the database. The algorithm is terminated when no further extensions that satisfy the minimum support condition are found. The main idea of the algorithm is scanning the database for frequent item sets, while on each following step pruning those items that are found to be infrequent. There are two very important steps in the candidate generation – the join and the prune step. In the first step, joining L_k with itself results in the generation of C_{k+1}. While in the prune step, if there is any k-item sets that is infrequent it is pruned because it cannot be a subset of the frequent (k+1) item set.

Ck – candidate item sets with size k.

Lk – frequent item sets with size k.

The Apriori algorithm can be represented in the following steps:

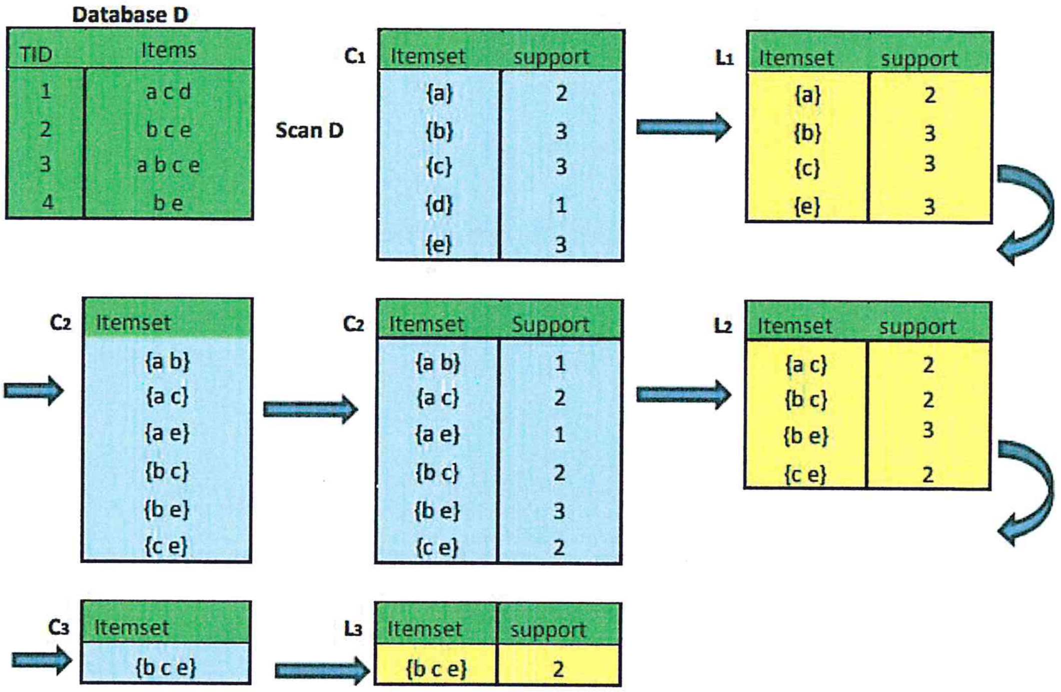
1. Find frequent items and put to the L_k (k=1).
2. Use L_k to generate a collection of candidate item sets C_{k+1} with size (k+1).
3. Scan the database to find which items in C_{k+1} are frequent and put them into L_{k+1}.
4. If L_{k+1} is not empty:

- K=k+1
- Go to step 2.

The example below shows how the Apriori works in a few simple steps. Let's suppose that a sample database of transactions consists of the following sets: {a, c, d}, {b, c, e}, {a, b, c, e}, {b, e}. Each letter corresponds to a certain product from the assortment. For example {a} is Iphone 5, {b} is Iphone 5 headphones.

On the first step, the algorithm counts up the frequencies of each item separately, also called supports. If we want to be sure that an item is frequent, we can predefine the minimum support level. In this case, the minimum support is 2. Therefore, four of the items are found to be frequent.

In the next step a list of all the 2-pairs of frequent items is generated. The already found infrequent items are excluded for further analysis. In order to find all possible two-item pairs, the Apriori algorithm prunes of the all possible combinations. At the last step, by connecting a frequent pair to a frequent single item a list of all the three-triplets of frequent items is generated. The algorithm ends at this step, because the pair of four items generated at the next step doesn't meet the required minimum support.



Database D

TID	Items
1	a c d
2	b c e
3	a b c e
4	b e

C1

Itemset	support
{a}	2
{b}	3
{c}	3
{d}	1
{e}	3

L1

Itemset	support
{a}	2
{b}	3
{c}	3
{e}	3

C2

Itemset
{a b}
{a c}
{a e}
{b c}
{b e}
{c e}

C2

Itemset	Support
{a b}	1
{a c}	2
{a e}	1
{b c}	2
{b e}	3
{c e}	2

L2

Itemset	support
{a c}	2
{b c}	2
{b e}	3
{c e}	2

C3

Itemset
{b c e}

L3

Itemset	support
{b c e}	2

Figure 4: Apriori Algorithm Architecture

4.4 Application of The Algorithm to Data

We have different kind of basket and product segmentation level. We will analyze the data step by step as below.

4.4.1 Looking Baskets Based on Payment Code

The item list dedicate based on payment code column and each basket detail could see what is include the basket or which items are bought together. Therefore, one of the necessary question what is the items in the data. We can use the Product Id, Catalog Id or

Category Id as item. Payment code item list which consist on Catalog Id's as item in Table2. We will analyze other items with payment code in this section.

Payment Code	Item1	Item2	Item3
68894865	909		
69630380	5246	5244	
77213152	7290	8766	8760
77213987	1575	5797	
77215761	5789	7290	5797

Table 2: Payment Code and Catalog Id Table Data Example

4.4.2 Create More Efficient Item list

The transaction data has empty value of some variables such as Catalog Id. If these measurements remove the data, item list size decreases. Hence, we may lose good information about the customer shopping habits. We think that what to do for resolving null data negative effects. The Catalog Id and Category Id kind of similar two column in the data. Catalog Id is an item identification number for a specific product, Category Id is kindly small group category of a product for example, a headphone. If we will see the rules that becomes Category Id, we know the category of the products but we cannot know the specific features. So, these table is seen Table 3.

Payment Code	Item1	Item2	Item3
72159859	5797	tc	
72161754	4168	taf	
72165878	5725	5511	tc
72168677	5244	taf	

72225167	5798	Taf	tah

Table 3: Payment Code and Catalog Id and Category Id Table Data Example

4.4.3 Looking Baskets Based on Member Id

Payment code is used for discovering the shopping baskets and it is used lots of analysis with different combinations of the algorithm. However, we think that member id is also used to discover consumer habits because the data is 6 months old. In addition, if we use member id with combination of Category Id and Catalog Id, we may find better information and better rules. The data table can be seen Table 4.

Member Id	Item1	Item2	Item3
3608	355	tc	taf
5378	tae1	tabc	tah
15217	8886	taf	tc
15518	7290	taf	taf
18509	7291	taf	

Table 4: Member Id and Catalog Id and Category Id Table Data Example

5. RESULTS

5.1 Results for Model No 1

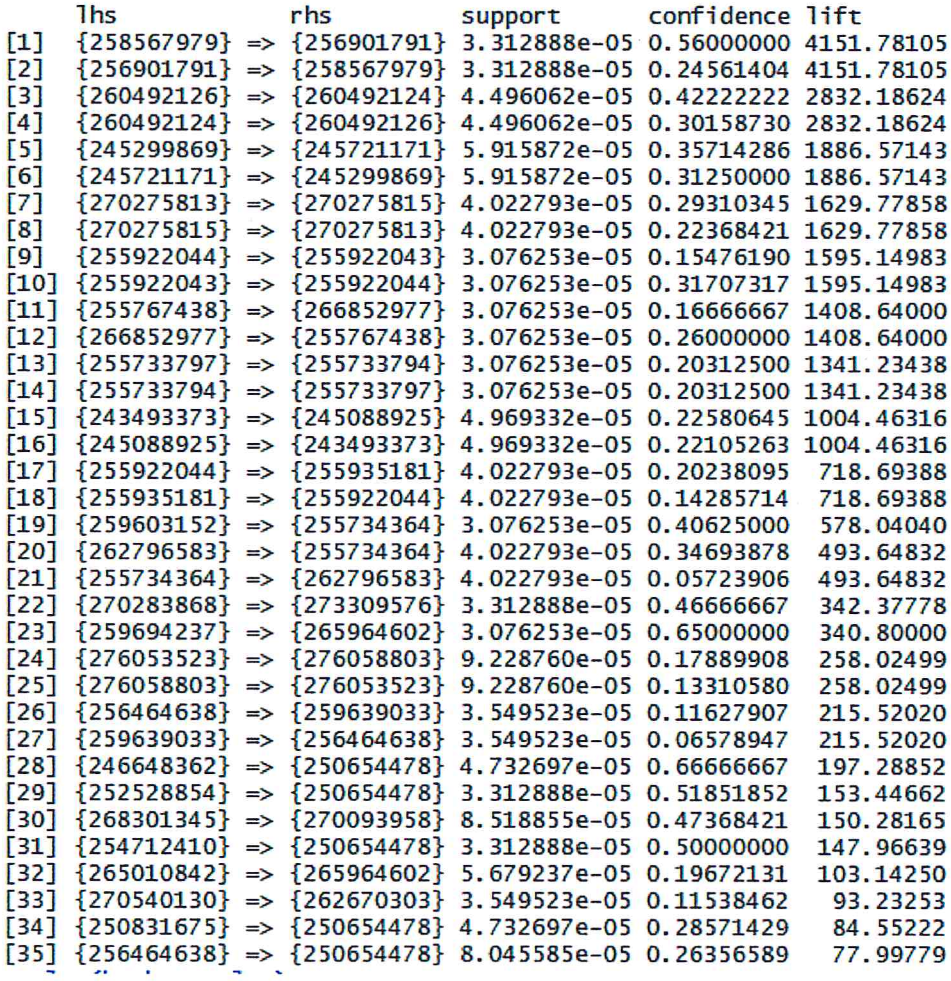
Model 1 consists of Payment Code and Product Id item list. The purpose of this model is to investigate which Product Id is bought in the specific basket. In other words we want to see which product is more likely to purchase to another product. The analysis will provide valuable information to the retailer so he can adjust promotional activities accordingly.

Rules with higher support and confidence values are called strong rules. If support is a big value and there is no rule created, we can decrease the support level to see the normal level rules. Then, we must decide for the confidence and minimum length of the rules.

Parameter	Value
Minlen	2
Support	0,00003
Confidence	0,05

Table 5: Model No 1 Parameters Table

If parameters are set as Table 5, the algorithm gives the output as Figure 4.



```
     lhs            rhs         support      confidence lift
[1]  {258567979} => {256901791} 3.312888e-05 0.56000000 4151.78105
[2]  {256901791} => {258567979} 3.312888e-05 0.24561404 4151.78105
[3]  {260492126} => {260492124} 4.496062e-05 0.42222222 2832.18624
[4]  {260492124} => {260492126} 4.496062e-05 0.30158730 2832.18624
[5]  {245299869} => {245721171} 5.915872e-05 0.35714286 1886.57143
[6]  {245721171} => {245299869} 5.915872e-05 0.31250000 1886.57143
[7]  {270275813} => {270275815} 4.022793e-05 0.29310345 1629.77858
[8]  {270275815} => {270275813} 4.022793e-05 0.22368421 1629.77858
[9]  {255922044} => {255922043} 3.076253e-05 0.15476190 1595.14983
[10] {255922043} => {255922044} 3.076253e-05 0.31707317 1595.14983
[11] {255767438} => {266852977} 3.076253e-05 0.16666667 1408.64000
[12] {266852977} => {255767438} 3.076253e-05 0.26000000 1408.64000
[13] {255733797} => {255733794} 3.076253e-05 0.20312500 1341.23438
[14] {255733794} => {255733797} 3.076253e-05 0.20312500 1341.23438
[15] {243493373} => {245088925} 4.969332e-05 0.22580645 1004.46316
[16] {245088925} => {243493373} 4.969332e-05 0.22105263 1004.46316
[17] {255922044} => {255935181} 4.022793e-05 0.20238095  718.69388
[18] {255935181} => {255922044} 4.022793e-05 0.14285714  718.69388
[19] {259603152} => {255734364} 3.076253e-05 0.40625000  578.04040
[20] {262796583} => {255734364} 4.022793e-05 0.34693878  493.64832
[21] {255734364} => {262796583} 4.022793e-05 0.05723906  493.64832
[22] {270283868} => {273309576} 3.312888e-05 0.46666667  342.37778
[23] {259694237} => {265964602} 3.076253e-05 0.65000000  340.80000
[24] {276053523} => {276058803} 9.228760e-05 0.17889908  258.02499
[25] {276058803} => {276053523} 9.228760e-05 0.13310580  258.02499
[26] {256464638} => {259639033} 3.549523e-05 0.11627907  215.52020
[27] {259639033} => {256464638} 3.549523e-05 0.06578947  215.52020
[28] {246648362} => {250654478} 4.732697e-05 0.66666667  197.28852
[29] {252528854} => {250654478} 3.312888e-05 0.51851852  153.44662
[30] {268301345} => {270093958} 8.518855e-05 0.47368421  150.28165
[31] {254712410} => {250654478} 3.312888e-05 0.50000000  147.96639
[32] {265010842} => {265964602} 5.679237e-05 0.19672131  103.14250
[33] {270540130} => {262670303} 3.549523e-05 0.11538462   93.23253
[34] {250831675} => {250654478} 4.732697e-05 0.28571429   84.55222
[35] {256464638} => {250654478} 8.045585e-05 0.26356589   77.99779
```

Figure 5: Output of Model No 1

Apriori algorithm implies that 258567979 and 256901791 items are frequently purchased together with 0,56 confidence. Only 35 rules generated as Figure 6.

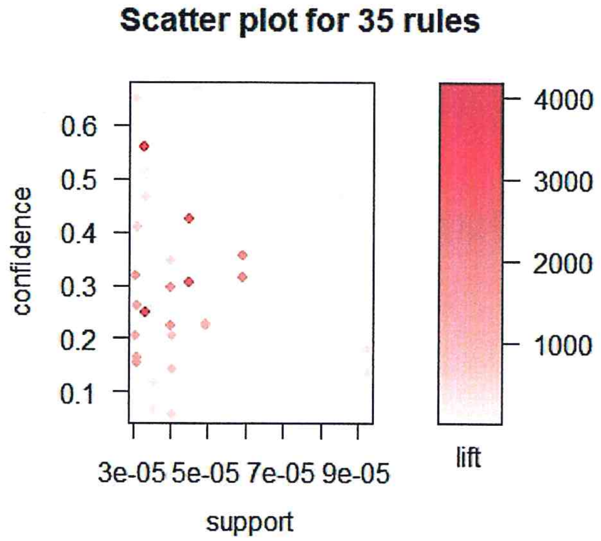


Figure 6: Model No 1 Scatterplot

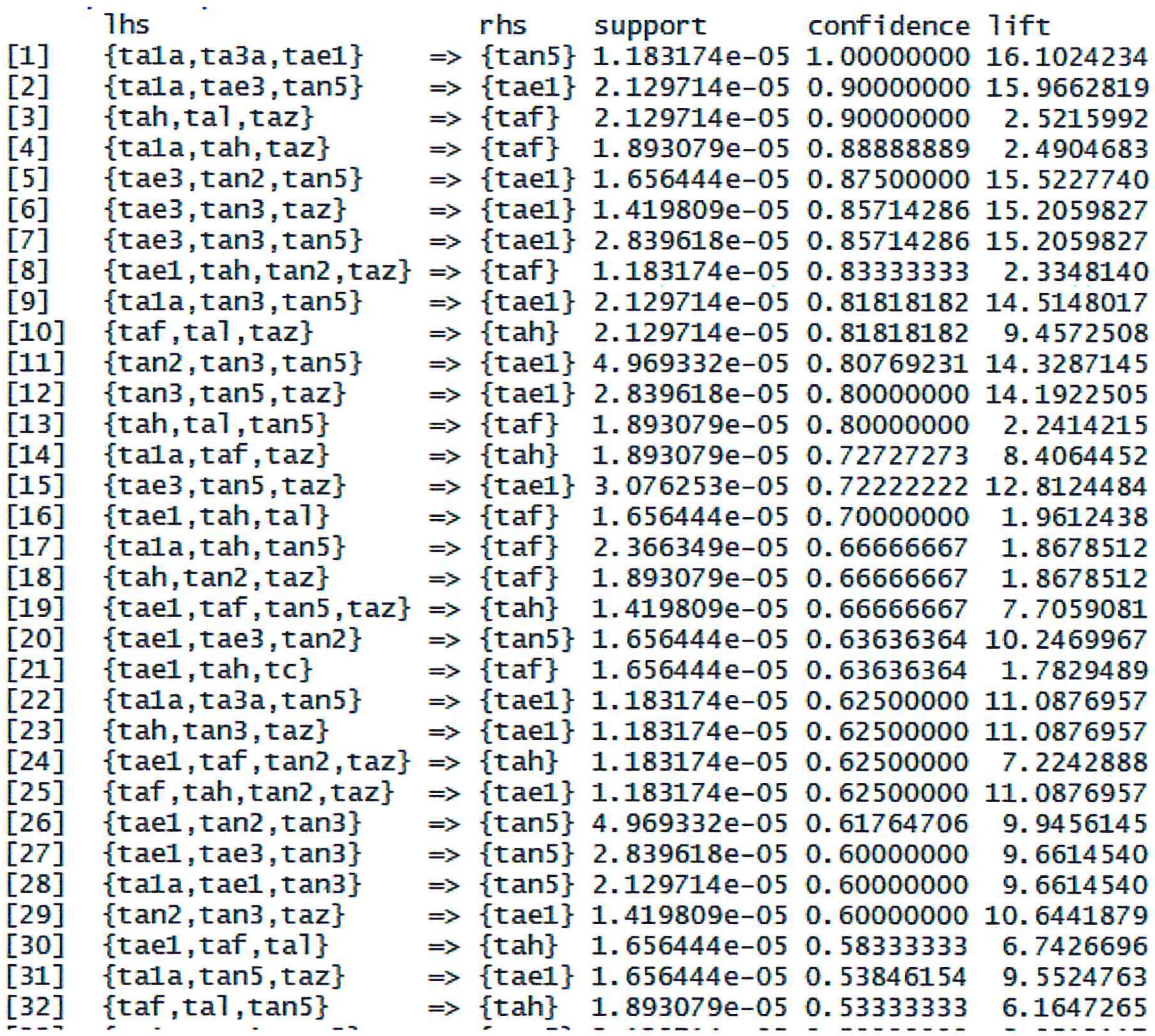
5.2 Results for Model No 2

Model 2 consists of Payment Code and Category Id item list. The purpose of this model is to investigate which category items are bought together. We do not know exactly what the product is, but we know the product category.

Parameter	Value
Minlen	4
Support	0,00001
Confidence	0,05

Table 6: Model No 2 Parameters Table

If parameters are set as Table 6, the algorithm gives the output as Figure 7.



```
     lhs                       rhs     support      confidence lift
[1]  {ta1a,ta3a,tae1}       => {tan5} 1.183174e-05 1.00000000 16.1024234
[2]  {ta1a,tae3,tan5}       => {tae1} 2.129714e-05 0.90000000 15.9662819
[3]  {tah,tal,taz}          => {taf}  2.129714e-05 0.90000000  2.5215992
[4]  {ta1a,tah,taz}         => {taf}  1.893079e-05 0.88888889  2.4904683
[5]  {tae3,tan2,tan5}       => {tae1} 1.656444e-05 0.87500000 15.5227740
[6]  {tae3,tan3,taz}        => {tae1} 1.419809e-05 0.85714286 15.2059827
[7]  {tae3,tan3,tan5}       => {tae1} 2.839618e-05 0.85714286 15.2059827
[8]  {tae1,tah,tan2,taz}    => {taf}  1.183174e-05 0.83333333  2.3348140
[9]  {ta1a,tan3,tan5}       => {tae1} 2.129714e-05 0.81818182 14.5148017
[10] {taf,tal,taz}          => {tah}  2.129714e-05 0.81818182  9.4572508
[11] {tan2,tan3,tan5}       => {tae1} 4.969332e-05 0.80769231 14.3287145
[12] {tan3,tan5,taz}        => {tae1} 2.839618e-05 0.80000000 14.1922505
[13] {tah,tal,tan5}         => {taf}  1.893079e-05 0.80000000  2.2414215
[14] {ta1a,taf,taz}         => {tah}  1.893079e-05 0.72727273  8.4064452
[15] {tae3,tan5,taz}        => {tae1} 3.076253e-05 0.72222222 12.8124484
[16] {tae1,tah,tal}         => {taf}  1.656444e-05 0.70000000  1.9612438
[17] {ta1a,tah,tan5}        => {taf}  2.366349e-05 0.66666667  1.8678512
[18] {tah,tan2,taz}         => {taf}  1.893079e-05 0.66666667  1.8678512
[19] {tae1,taf,tan5,taz}    => {tah}  1.419809e-05 0.66666667  7.7059081
[20] {tae1,tae3,tan2}       => {tan5} 1.656444e-05 0.63636364 10.2469967
[21] {tae1,tah,tc}          => {taf}  1.656444e-05 0.63636364  1.7829489
[22] {ta1a,ta3a,tan5}       => {tae1} 1.183174e-05 0.62500000 11.0876957
[23] {tah,tan3,taz}         => {tae1} 1.183174e-05 0.62500000 11.0876957
[24] {tae1,taf,tan2,taz}    => {tah}  1.183174e-05 0.62500000  7.2242888
[25] {taf,tah,tan2,taz}     => {tae1} 1.183174e-05 0.62500000 11.0876957
[26] {tae1,tan2,tan3}       => {tan5} 4.969332e-05 0.61764706  9.9456145
[27] {tae1,tae3,tan3}       => {tan5} 2.839618e-05 0.60000000  9.6614540
[28] {ta1a,tae1,tan3}       => {tan5} 2.129714e-05 0.60000000  9.6614540
[29] {tan2,tan3,taz}        => {tae1} 1.419809e-05 0.60000000 10.6441879
[30] {tae1,taf,tal}         => {tah}  1.656444e-05 0.58333333  6.7426696
[31] {ta1a,tan5,taz}        => {tae1} 1.656444e-05 0.53846154  9.5524763
[32] {taf,tal,tan5}         => {tah}  1.893079e-05 0.53333333  6.1647265
```

Figure 7: Output of the Model No 2

Apriori algorithm implies that if the consumer buys ta1a, ta3a and tae1 category items in the one shopping basket, he/she may also buy a product in the tan5 category with %100 confidence. Only 162 rules generated as Figure 8.

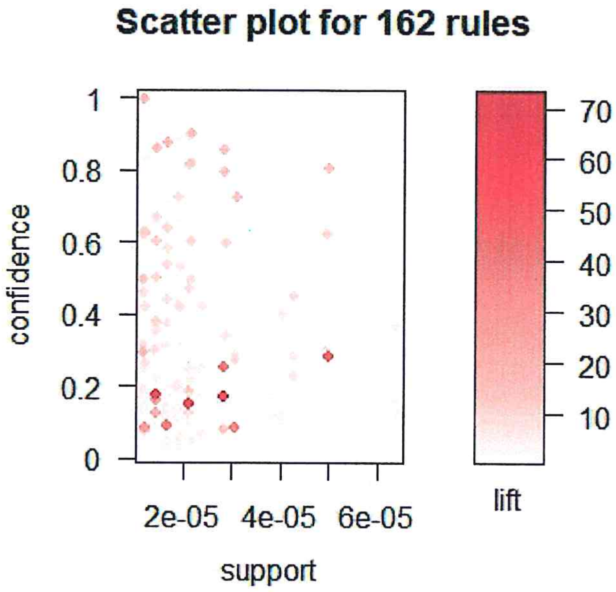


Figure 8: Model No 2 Scatterplot

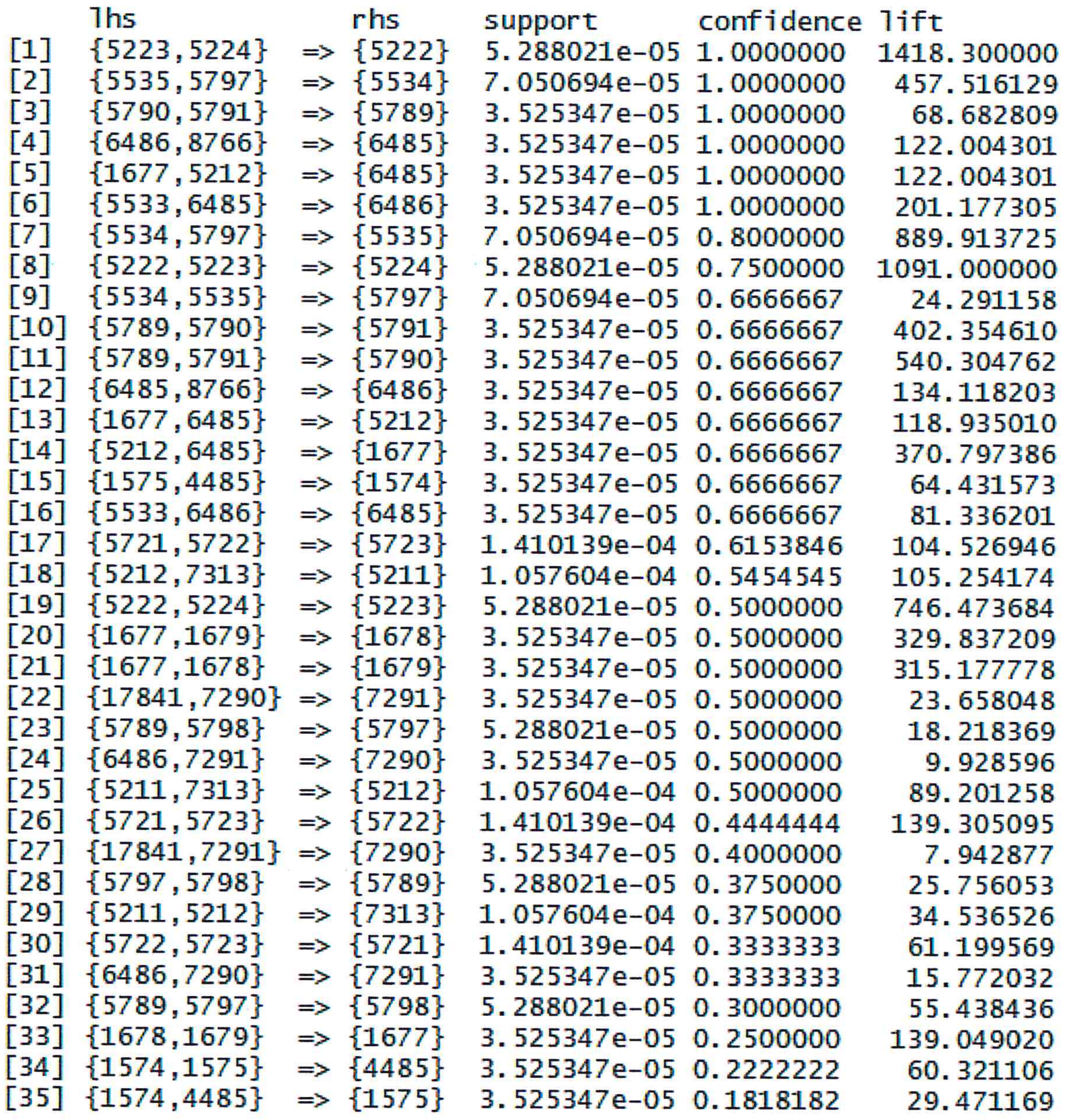
5.3 Results for Model No 3

Model 3 consists of Payment Code and Catalog Id item list. The purpose of this model is to investigate which catalog items are bought together. We do not know exactly what is the product, but we know the product catalog. We also remove the null value of the data before creating the item list.

Parameter	Value
Minlen	3
Support	0,00002
Confidence	0,05

Table 7: Model No 3 Parameters Table

If parameters are set as Table 7, the algorithm gives the output as Figure 9.



```
     lhs             rhs     support      confidence lift
[1]  {5223,5224}  => {5222}  5.288021e-05 1.0000000  1418.300000
[2]  {5535,5797}  => {5534}  7.050694e-05 1.0000000   457.516129
[3]  {5790,5791}  => {5789}  3.525347e-05 1.0000000    68.682809
[4]  {6486,8766}  => {6485}  3.525347e-05 1.0000000   122.004301
[5]  {1677,5212}  => {6485}  3.525347e-05 1.0000000   122.004301
[6]  {5533,6485}  => {6486}  3.525347e-05 1.0000000   201.177305
[7]  {5534,5797}  => {5535}  7.050694e-05 0.8000000   889.913725
[8]  {5222,5223}  => {5224}  5.288021e-05 0.7500000  1091.000000
[9]  {5534,5535}  => {5797}  7.050694e-05 0.6666667    24.291158
[10] {5789,5790}  => {5791}  3.525347e-05 0.6666667   402.354610
[11] {5789,5791}  => {5790}  3.525347e-05 0.6666667   540.304762
[12] {6485,8766}  => {6486}  3.525347e-05 0.6666667   134.118203
[13] {1677,6485}  => {5212}  3.525347e-05 0.6666667   118.935010
[14] {5212,6485}  => {1677}  3.525347e-05 0.6666667   370.797386
[15] {1575,4485}  => {1574}  3.525347e-05 0.6666667    64.431573
[16] {5533,6486}  => {6485}  3.525347e-05 0.6666667    81.336201
[17] {5721,5722}  => {5723}  1.410139e-04 0.6153846   104.526946
[18] {5212,7313}  => {5211}  1.057604e-04 0.5454545   105.254174
[19] {5222,5224}  => {5223}  5.288021e-05 0.5000000   746.473684
[20] {1677,1679}  => {1678}  3.525347e-05 0.5000000   329.837209
[21] {1677,1678}  => {1679}  3.525347e-05 0.5000000   315.177778
[22] {17841,7290} => {7291}  3.525347e-05 0.5000000    23.658048
[23] {5789,5798}  => {5797}  5.288021e-05 0.5000000    18.218369
[24] {6486,7291}  => {7290}  3.525347e-05 0.5000000     9.928596
[25] {5211,7313}  => {5212}  1.057604e-04 0.5000000    89.201258
[26] {5721,5723}  => {5722}  1.410139e-04 0.4444444   139.305095
[27] {17841,7291} => {7290}  3.525347e-05 0.4000000     7.942877
[28] {5797,5798}  => {5789}  5.288021e-05 0.3750000    25.756053
[29] {5211,5212}  => {7313}  1.057604e-04 0.3750000    34.536526
[30] {5722,5723}  => {5721}  1.410139e-04 0.3333333    61.199569
[31] {6486,7290}  => {7291}  3.525347e-05 0.3333333    15.772032
[32] {5789,5797}  => {5798}  5.288021e-05 0.3000000    55.438436
[33] {1678,1679}  => {1677}  3.525347e-05 0.2500000   139.049020
[34] {1574,1575}  => {4485}  3.525347e-05 0.2222222    60.321106
[35] {1574,4485}  => {1575}  3.525347e-05 0.1818182    29.471169
```

Figure 9: Output of the Model No 3

Apriori algorithm implies that if the consumer buys 5223 and 52241 catalog items in the one shopping basket, he/she may also buys a product in the 5222 catalog id with %100 confidence. Only 39 rules generated as Figure 10.

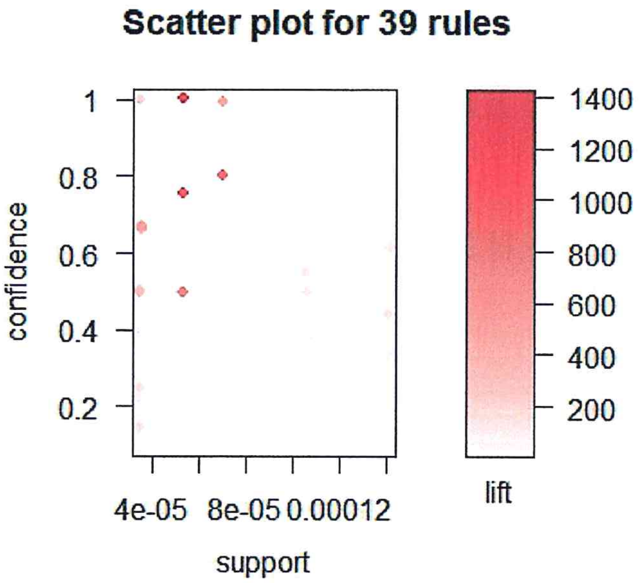


Figure 10: Model No 3 Scatterplot

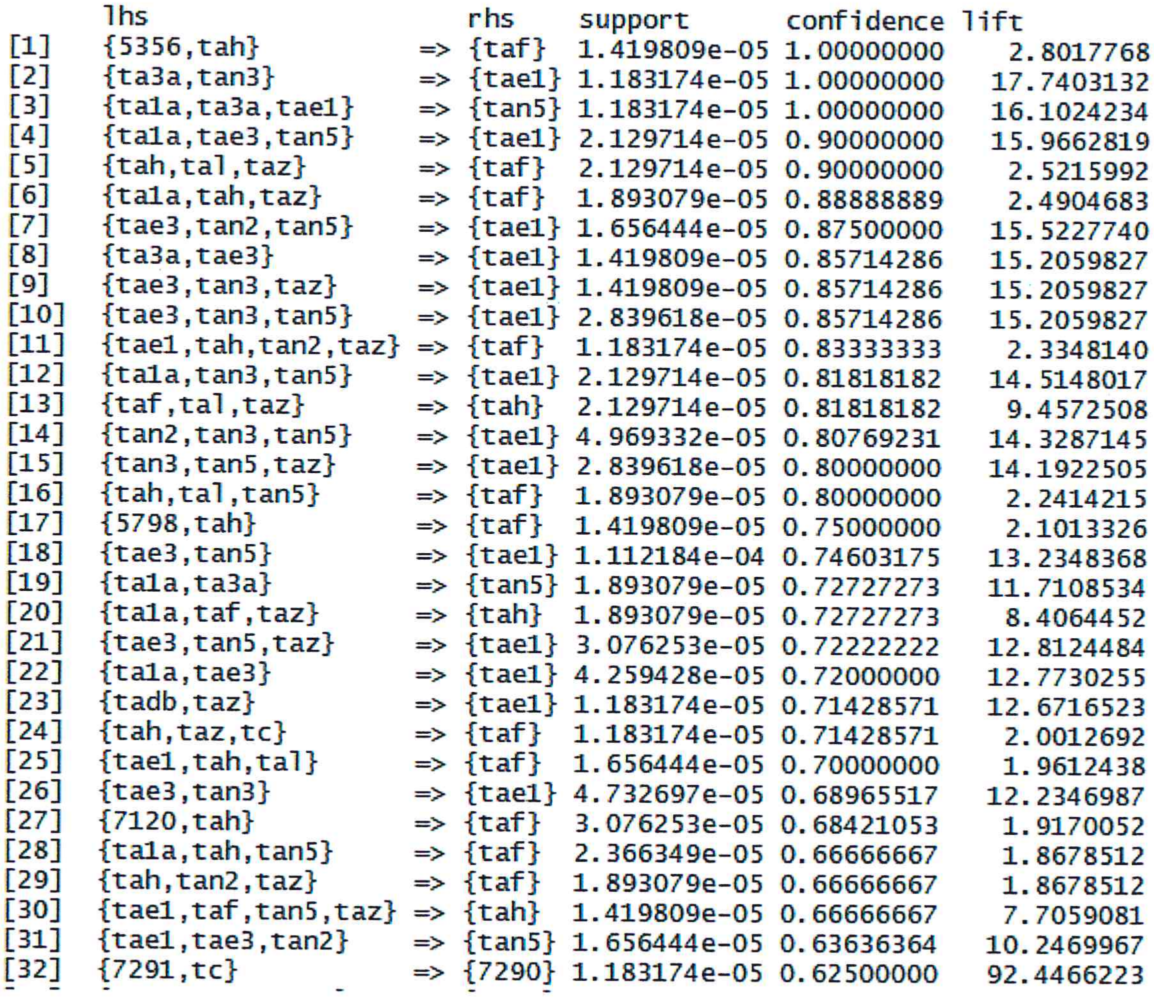
5.4 Results for Model No 4

Model 4 consists of Payment Code and combination of Catalog Id and Category Id item list. The purpose of this model is to investigate which items are bought together. We do not know exactly product that category id known, but we know product that has catalog id.

Parameter	Value
Minlen	3
Support	0,00001
Confidence	0,05

Table 8: Model No 4 Parameters

If parameters are set as Table 8, the algorithm gives the output as Figure 11.



```
     lhs                          rhs     support      confidence lift
[1]  {5356,tah}                => {taf}   1.419809e-05 1.00000000  2.8017768
[2]  {ta3a,tan3}               => {tae1}  1.183174e-05 1.00000000 17.7403132
[3]  {ta1a,ta3a,tae1}          => {tan5}  1.183174e-05 1.00000000 16.1024234
[4]  {ta1a,tae3,tan5}          => {tae1}  2.129714e-05 0.90000000 15.9662819
[5]  {tah,tal,taz}             => {taf}   2.129714e-05 0.90000000  2.5215992
[6]  {ta1a,tah,taz}            => {taf}   1.893079e-05 0.88888889  2.4904683
[7]  {tae3,tan2,tan5}          => {tae1}  1.656444e-05 0.87500000 15.5227740
[8]  {ta3a,tae3}               => {tae1}  1.419809e-05 0.85714286 15.2059827
[9]  {tae3,tan3,taz}           => {tae1}  1.419809e-05 0.85714286 15.2059827
[10] {tae3,tan3,tan5}          => {tae1}  2.839618e-05 0.85714286 15.2059827
[11] {tae1,tah,tan2,taz}       => {taf}   1.183174e-05 0.83333333  2.3348140
[12] {ta1a,tan3,tan5}          => {tae1}  2.129714e-05 0.81818182 14.5148017
[13] {taf,tal,taz}             => {tah}   2.129714e-05 0.81818182  9.4572508
[14] {tan2,tan3,tan5}          => {tae1}  4.969332e-05 0.80769231 14.3287145
[15] {tan3,tan5,taz}           => {tae1}  2.839618e-05 0.80000000 14.1922505
[16] {tah,tal,tan5}            => {taf}   1.893079e-05 0.80000000  2.2414215
[17] {5798,tah}                => {taf}   1.419809e-05 0.75000000  2.1013326
[18] {tae3,tan5}               => {tae1}  1.112184e-04 0.74603175 13.2348368
[19] {ta1a,ta3a}               => {tan5}  1.893079e-05 0.72727273 11.7108534
[20] {ta1a,taf,taz}            => {tah}   1.893079e-05 0.72727273  8.4064452
[21] {tae3,tan5,taz}           => {tae1}  3.076253e-05 0.72222222 12.8124484
[22] {ta1a,tae3}               => {tae1}  4.259428e-05 0.72000000 12.7730255
[23] {tadb,taz}                => {tae1}  1.183174e-05 0.71428571 12.6716523
[24] {tah,taz,tc}              => {taf}   1.183174e-05 0.71428571  2.0012692
[25] {tae1,tah,tal}            => {taf}   1.656444e-05 0.70000000  1.9612438
[26] {tae3,tan3}               => {tae1}  4.732697e-05 0.68965517 12.2346987
[27] {7120,tah}                => {taf}   3.076253e-05 0.68421053  1.9170052
[28] {ta1a,tah,tan5}           => {taf}   2.366349e-05 0.66666667  1.8678512
[29] {tah,tan2,taz}            => {taf}   1.893079e-05 0.66666667  1.8678512
[30] {tae1,taf,tan5,taz}       => {tah}   1.419809e-05 0.66666667  7.7059081
[31] {tae1,tae3,tan2}          => {tan5}  1.656444e-05 0.63636364 10.2469967
[32] {7291,tc}                 => {7290}  1.183174e-05 0.62500000 92.4466223
```

Figure 11: Output of the Model No 4

Apriori algorithm implies that if the consumer buys that catalog id is 5356 and category is tah in the one shopping basket, he/she may also buys a product in the taf category with %100 confidence. Only 446 rules generated as Figure 12.

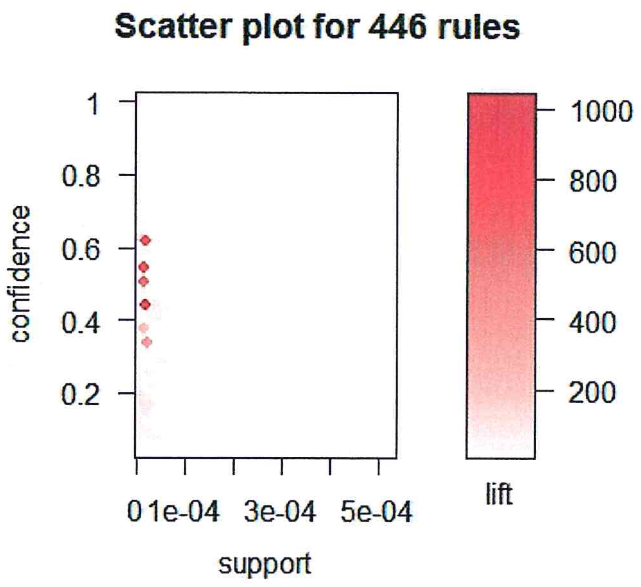


Figure 12: Model No 4 Scatterplot

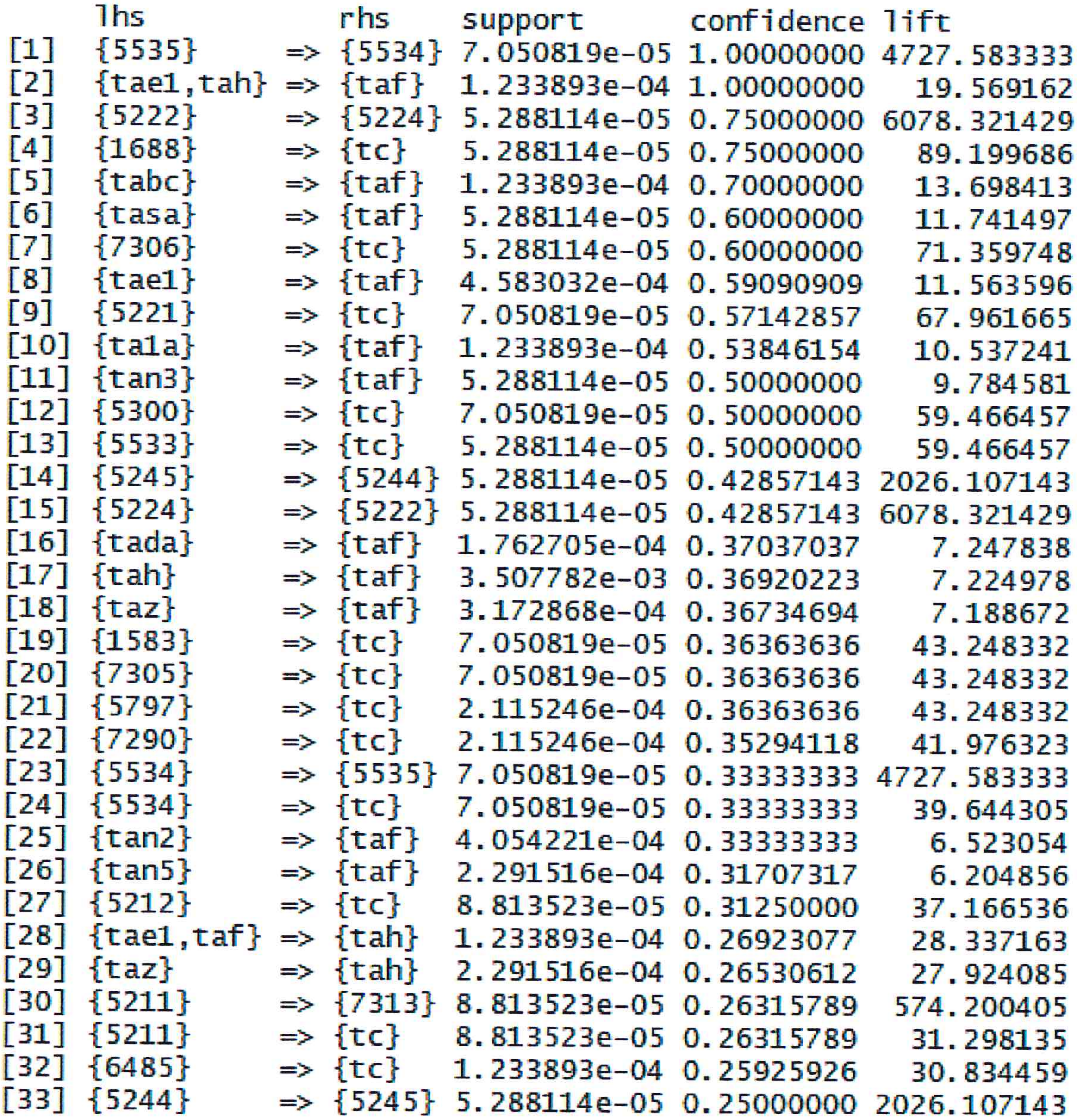
5.5 Results for Model No 5

Model 5 consists of Payment Code and combination of Catalog Id and Category Id item list which every row must have at least one Catalog Id. The purpose of this model is to investigate which items are bought together. We do not know exactly product that category id known, but we know product that has catalog id.

Parameter	Value
Minlen	2
Support	0,00005
Confidence	0,05

Table 9: Model No 5 Parameters

If parameters are set as Table 9, the algorithm gives the output as Figure 13.



```
     lhs            rhs    support      confidence lift
[1]  {5535}      => {5534} 7.050819e-05 1.00000000 4727.583333
[2]  {tae1,tah} => {taf}  1.233893e-04 1.00000000   19.569162
[3]  {5222}      => {5224} 5.288114e-05 0.75000000 6078.321429
[4]  {1688}      => {tc}   5.288114e-05 0.75000000   89.199686
[5]  {tabc}      => {taf}  1.233893e-04 0.70000000   13.698413
[6]  {tasa}      => {taf}  5.288114e-05 0.60000000   11.741497
[7]  {7306}      => {tc}   5.288114e-05 0.60000000   71.359748
[8]  {tae1}      => {taf}  4.583032e-04 0.59090909   11.563596
[9]  {5221}      => {tc}   7.050819e-05 0.57142857   67.961665
[10] {ta1a}      => {taf}  1.233893e-04 0.53846154   10.537241
[11] {tan3}      => {taf}  5.288114e-05 0.50000000    9.784581
[12] {5300}      => {tc}   7.050819e-05 0.50000000   59.466457
[13] {5533}      => {tc}   5.288114e-05 0.50000000   59.466457
[14] {5245}      => {5244} 5.288114e-05 0.42857143 2026.107143
[15] {5224}      => {5222} 5.288114e-05 0.42857143 6078.321429
[16] {tada}      => {taf}  1.762705e-04 0.37037037    7.247838
[17] {tah}       => {taf}  3.507782e-03 0.36920223    7.224978
[18] {taz}       => {taf}  3.172868e-04 0.36734694    7.188672
[19] {1583}      => {tc}   7.050819e-05 0.36363636   43.248332
[20] {7305}      => {tc}   7.050819e-05 0.36363636   43.248332
[21] {5797}      => {tc}   2.115246e-04 0.36363636   43.248332
[22] {7290}      => {tc}   2.115246e-04 0.35294118   41.976323
[23] {5534}      => {5535} 7.050819e-05 0.33333333 4727.583333
[24] {5534}      => {tc}   7.050819e-05 0.33333333   39.644305
[25] {tan2}      => {taf}  4.054221e-04 0.33333333    6.523054
[26] {tan5}      => {taf}  2.291516e-04 0.31707317    6.204856
[27] {5212}      => {tc}   8.813523e-05 0.31250000   37.166536
[28] {tae1,taf} => {tah}  1.233893e-04 0.26923077   28.337163
[29] {taz}       => {tah}  2.291516e-04 0.26530612   27.924085
[30] {5211}      => {7313} 8.813523e-05 0.26315789  574.200405
[31] {5211}      => {tc}   8.813523e-05 0.26315789   31.298135
[32] {6485}      => {tc}   1.233893e-04 0.25925926   30.834459
[33] {5244}      => {5245} 5.288114e-05 0.25000000 2026.107143
```

Figure 13: Output of the Model No 5

Apriori algorithm implies that if the consumer buys that catalog id is 5535, he/she may also buys that catalog id is 5534 with %100 confidence. Only 66 rules generated as Figure 14.

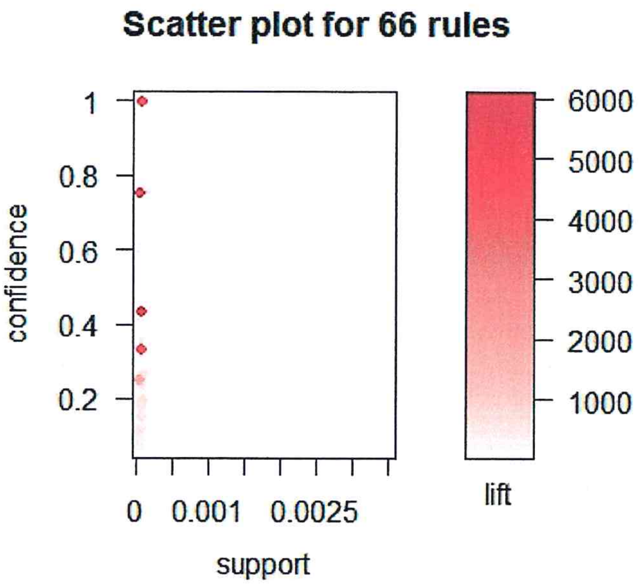


Figure 14: Model No 5 Scatterplot

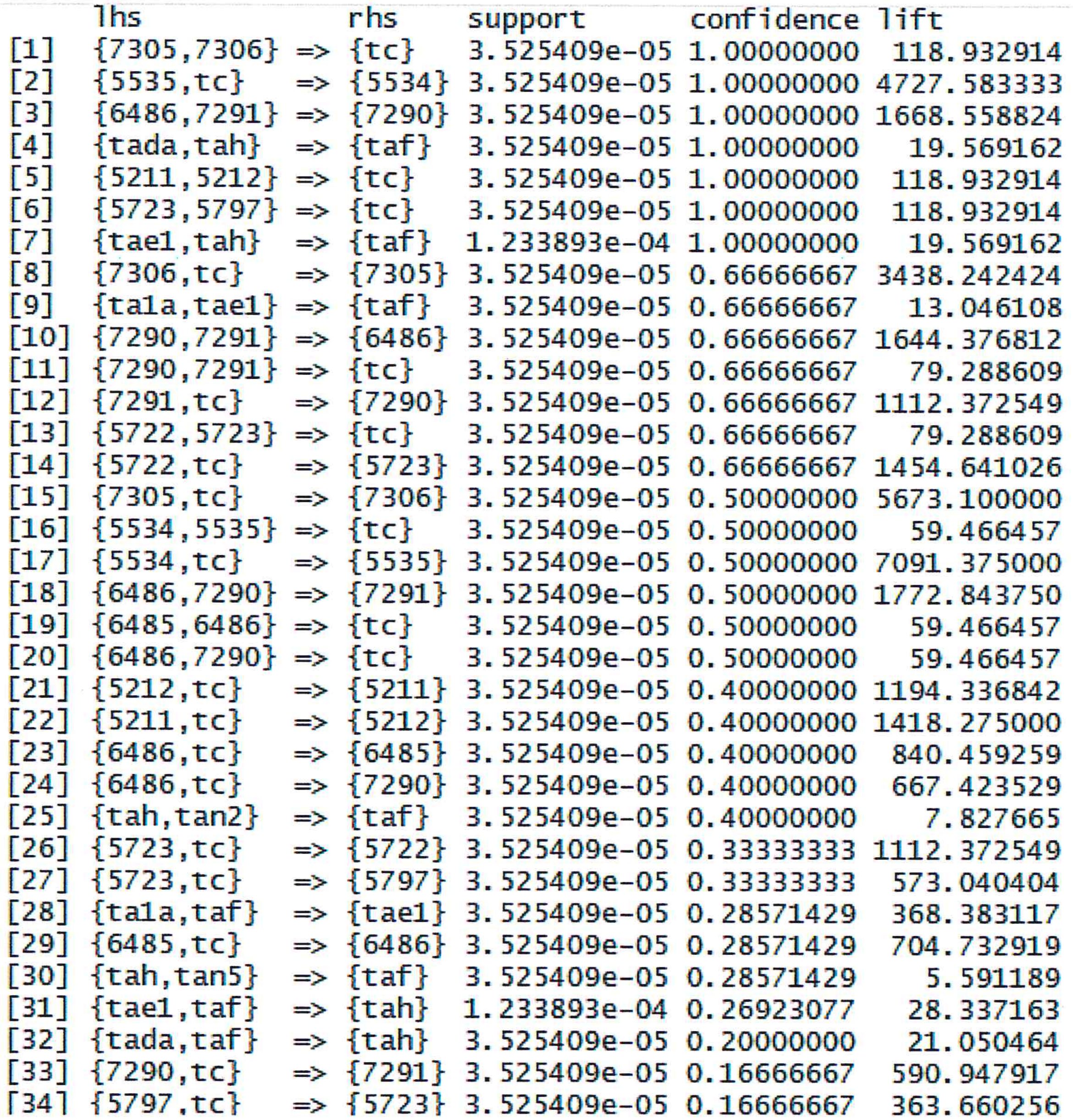
5.6 Results for Model No 6

Model 6 consists of Member Id and combination of Catalog Id and Category Id item list which every row must have at least one Catalog Id. The purpose of this model is to investigate which items are bought together. We do not know exactly product that category id known, but we know product that has catalog id.

Parameter	Value
Minlen	3
Support	0,00003
Confidence	0,05

Table 10: Model No 6 Parameters

If parameters are set as Table 10, the algorithm gives the output as Figure 15.



```
     lhs            rhs      support      confidence  lift
[1]  {7305,7306} => {tc}     3.525409e-05 1.00000000  118.932914
[2]  {5535,tc}   => {5534}   3.525409e-05 1.00000000 4727.583333
[3]  {6486,7291} => {7290}   3.525409e-05 1.00000000 1668.558824
[4]  {tada,tah}  => {taf}    3.525409e-05 1.00000000   19.569162
[5]  {5211,5212} => {tc}     3.525409e-05 1.00000000  118.932914
[6]  {5723,5797} => {tc}     3.525409e-05 1.00000000  118.932914
[7]  {tae1,tah}  => {taf}    1.233893e-04 1.00000000   19.569162
[8]  {7306,tc}   => {7305}   3.525409e-05 0.66666667 3438.242424
[9]  {ta1a,tae1} => {taf}    3.525409e-05 0.66666667   13.046108
[10] {7290,7291} => {6486}   3.525409e-05 0.66666667 1644.376812
[11] {7290,7291} => {tc}     3.525409e-05 0.66666667   79.288609
[12] {7291,tc}   => {7290}   3.525409e-05 0.66666667 1112.372549
[13] {5722,5723} => {tc}     3.525409e-05 0.66666667   79.288609
[14] {5722,tc}   => {5723}   3.525409e-05 0.66666667 1454.641026
[15] {7305,tc}   => {7306}   3.525409e-05 0.50000000 5673.100000
[16] {5534,5535} => {tc}     3.525409e-05 0.50000000   59.466457
[17] {5534,tc}   => {5535}   3.525409e-05 0.50000000 7091.375000
[18] {6486,7290} => {7291}   3.525409e-05 0.50000000 1772.843750
[19] {6485,6486} => {tc}     3.525409e-05 0.50000000   59.466457
[20] {6486,7290} => {tc}     3.525409e-05 0.50000000   59.466457
[21] {5212,tc}   => {5211}   3.525409e-05 0.40000000 1194.336842
[22] {5211,tc}   => {5212}   3.525409e-05 0.40000000 1418.275000
[23] {6486,tc}   => {6485}   3.525409e-05 0.40000000  840.459259
[24] {6486,tc}   => {7290}   3.525409e-05 0.40000000  667.423529
[25] {tah,tan2}  => {taf}    3.525409e-05 0.40000000    7.827665
[26] {5723,tc}   => {5722}   3.525409e-05 0.33333333 1112.372549
[27] {5723,tc}   => {5797}   3.525409e-05 0.33333333  573.040404
[28] {ta1a,taf}  => {tae1}   3.525409e-05 0.28571429  368.383117
[29] {6485,tc}   => {6486}   3.525409e-05 0.28571429  704.732919
[30] {tah,tan5}  => {taf}    3.525409e-05 0.28571429    5.591189
[31] {tae1,taf}  => {tah}    1.233893e-04 0.26923077   28.337163
[32] {tada,taf}  => {tah}    3.525409e-05 0.20000000   21.050464
[33] {7290,tc}   => {7291}   3.525409e-05 0.16666667  590.947917
[34] {5797,tc}   => {5723}   3.525409e-05 0.16666667  363.660256
```

Figure 15: Output of Model No 6

Apriori algorithm implies that if the consumer buys 7305 and 7306 catalog items in the one shopping basket, he/she may also buys a product in the tc category with %100 confidence. Only 40 rules generated as Figure 16.

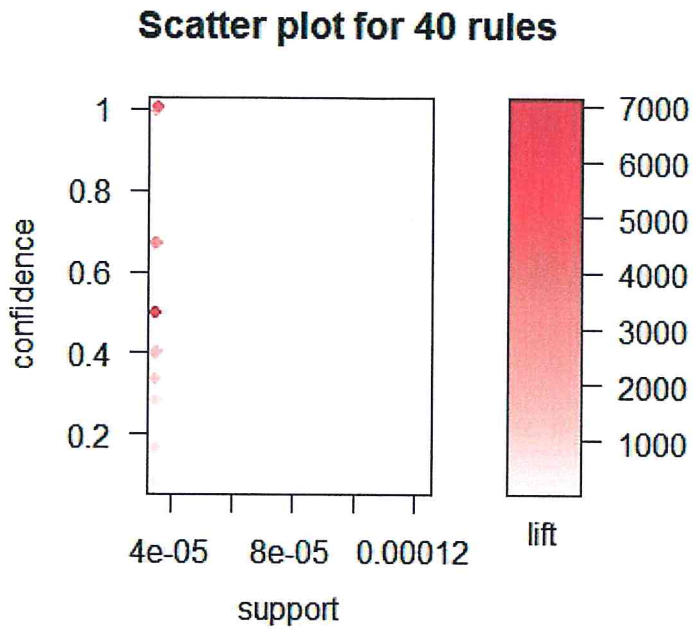


Figure 16: Model No 6 Scatterplot

6. SOCIAL AND ETHICAL APECTS

Mining into big data provides companies with a unique window into what is happening with business so that they can implement strategies efficiently. Obscure patterns can be discovered using market basket analysis which can help for planning more effective marketing efforts. It can be used not only for cross-sale and up-sale campaigns, but for managing better inventory control and satisfying shoppers' needs. Almost all departments of a company can benefit from a single analysis not only the high levels of management but also store operations, Merchandising and Advertising and Promotion departments.

Decision making and understanding the behavior of the customer has become vital and challenging problem for organizations to sustain their position in the competitive markets. Technological innovation shave paved breakthrough in faster processing of queries and sub-second response time. Data mining tools have become surest weapon for analyzing huge amount of data and breakthrough in making correct decisions. The objective of this project report is to analyze the huge amount of data thereby exploiting the consumer behavior and make the correct decision leading to competitive edge over rivals. Experimental analysis has been done employing association rules using Market Basket Analysis to prove its worth over the conventional methodologies.

7. CONTRIBUTION

The contributions of this project are that products purchased in bundles of 2, 3 and 4 were found in the transaction data. Moreover, Association Rules were generated more than one variable supporting probabilities and importance. Building up on previous researches by using established methods for mining association rules allowed for discovering useful information for the retailer. Evaluating probabilities of a basket membership depending on the category, because it provides the companies with better understanding of consumers' needs and suggests action for advertising.

APPENDIX A: RESULTS OF MODEL NO 1

The best 10 rules of model no 1 methodology are shown in Figure 17. Figure 18 is show us most 5 frequent items for the algorithm.

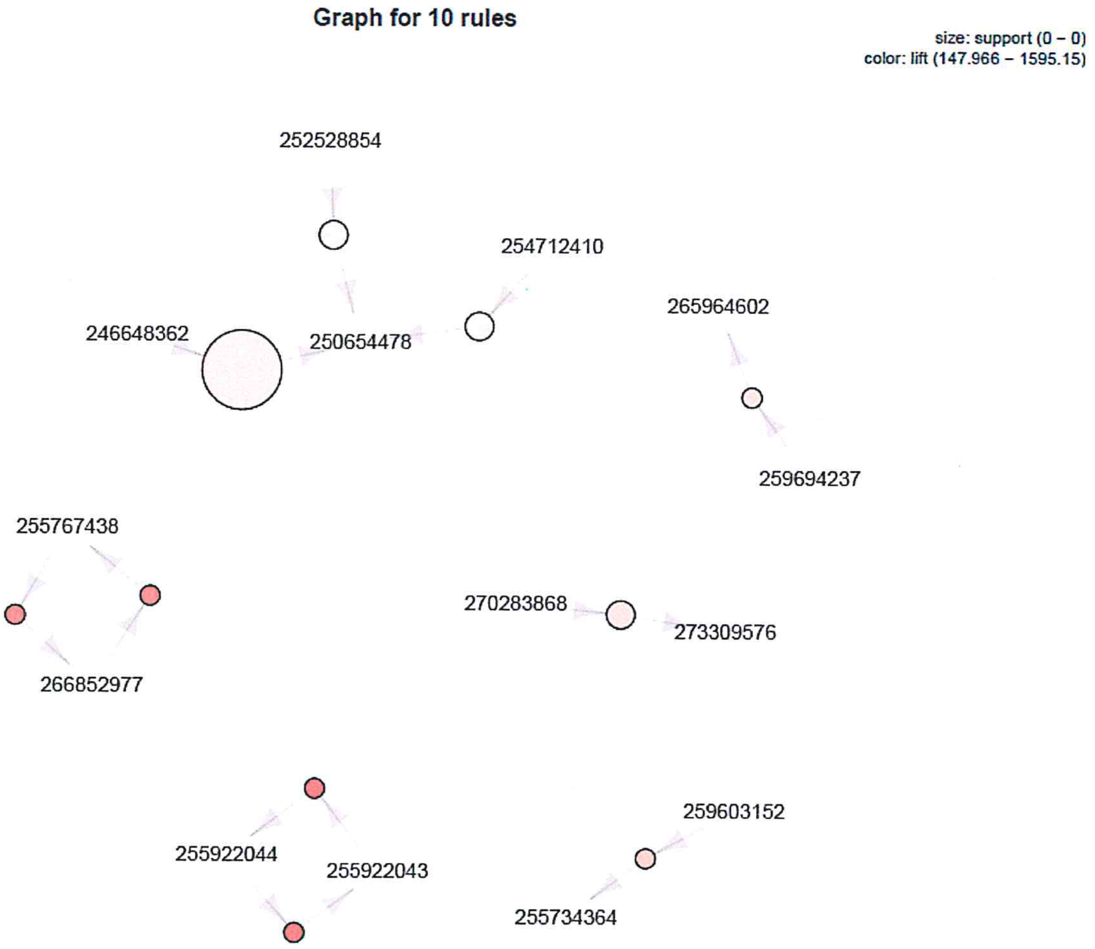


Figure 17: Best 10 Rules for Model No 1

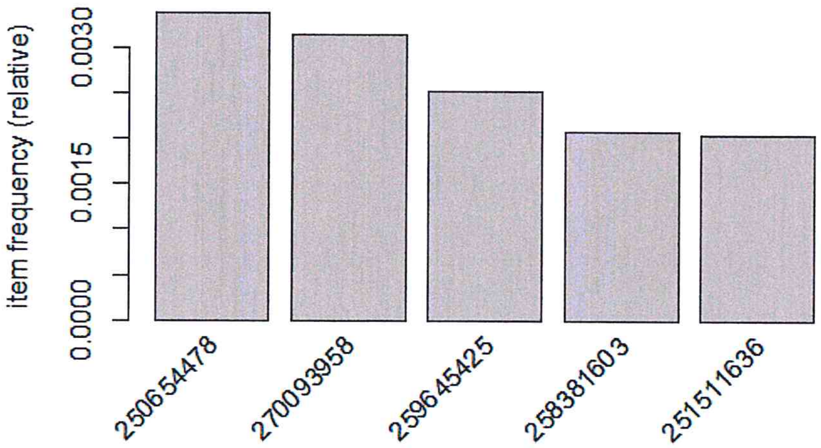


Figure 18: Most Frequent Items for Model No 1

APPENDIX B: RESULTS OF MODEL NO 2

The best 10 rules of model no 2 methodology are shown in Figure 19. Figure 20 is show us most 5 frequent items for the algorithm.

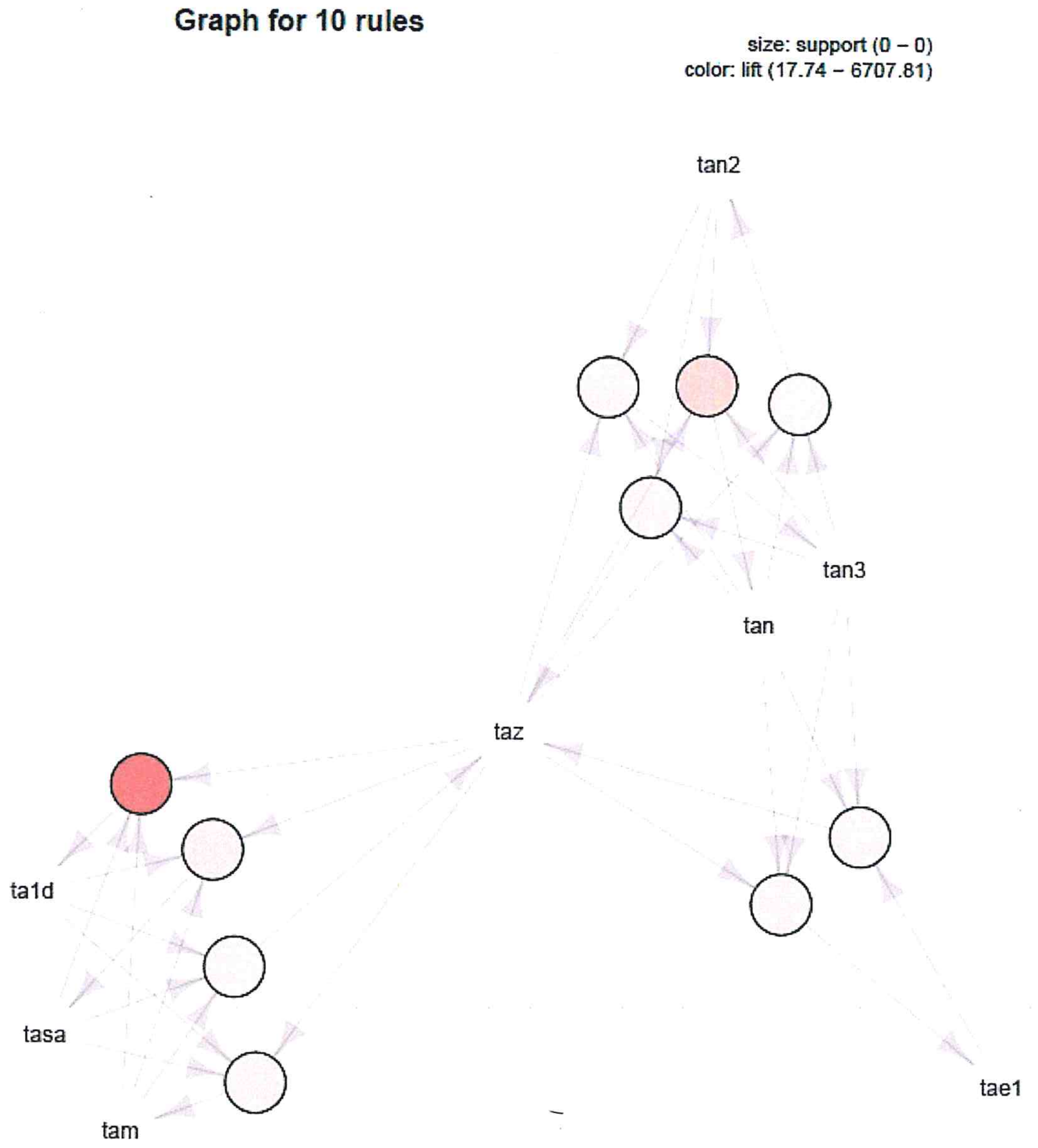


Figure 19: Best 10 Rules for Model No 2

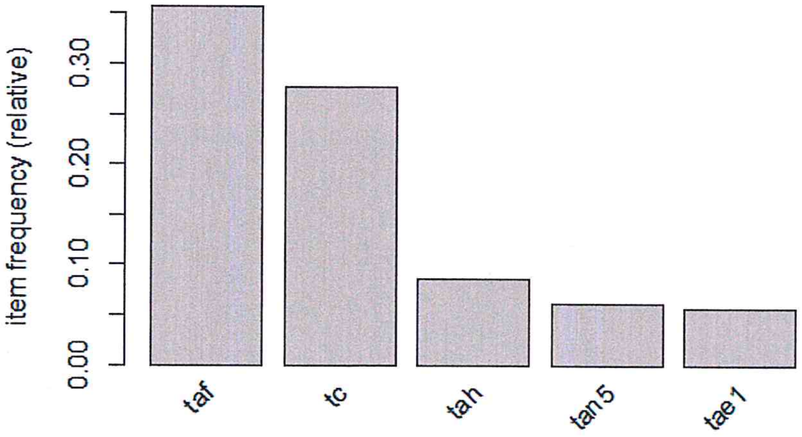


Figure 20: Most Frequent Items for Model No 2

APPENDIX C: RESULTS OF MODEL NO 3

The best 10 rules of model no 3 methodology are shown in Figure 21. Figure 22 is show us most 5 frequent items for the algorithm.

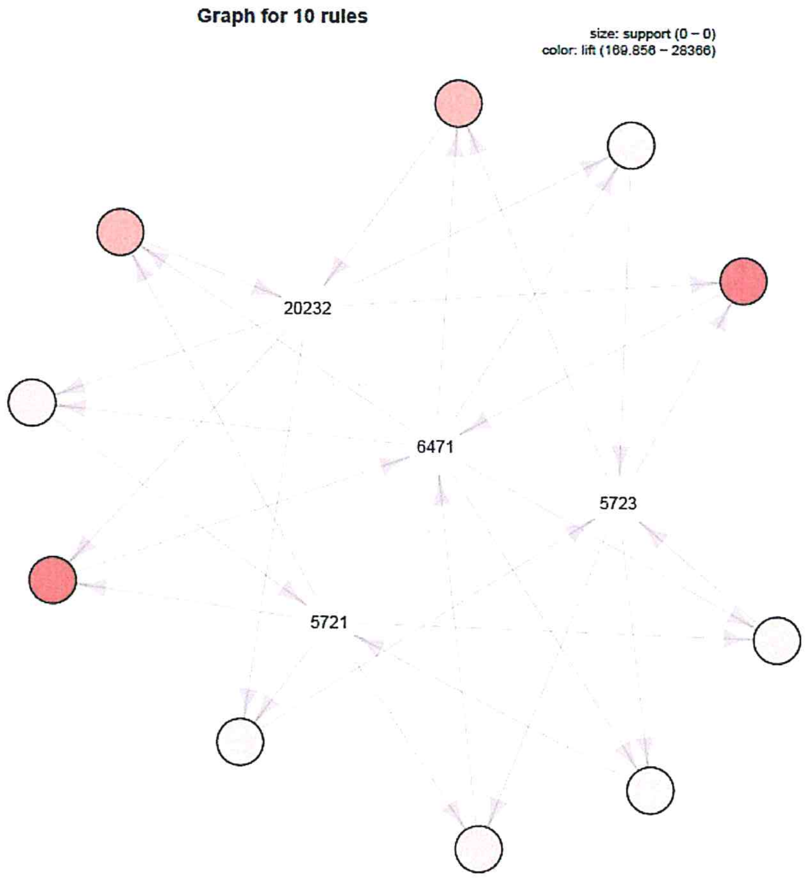


Figure 21: Best 10 Rules for Model No 3

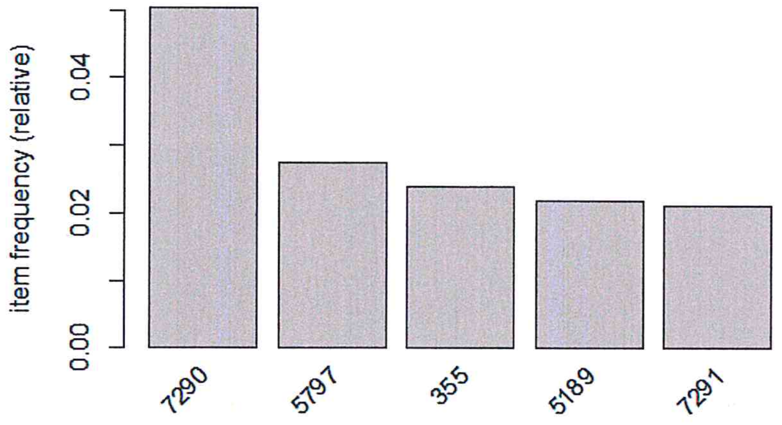


Figure 22: Most Frequent Items for Model No 3

APPENDIX D: RESULTS OF MODEL NO 4

The best 10 rules of model no 4 methodology are shown in Figure 23. Figure 24 is show us most 5 frequent items for the algorithm.

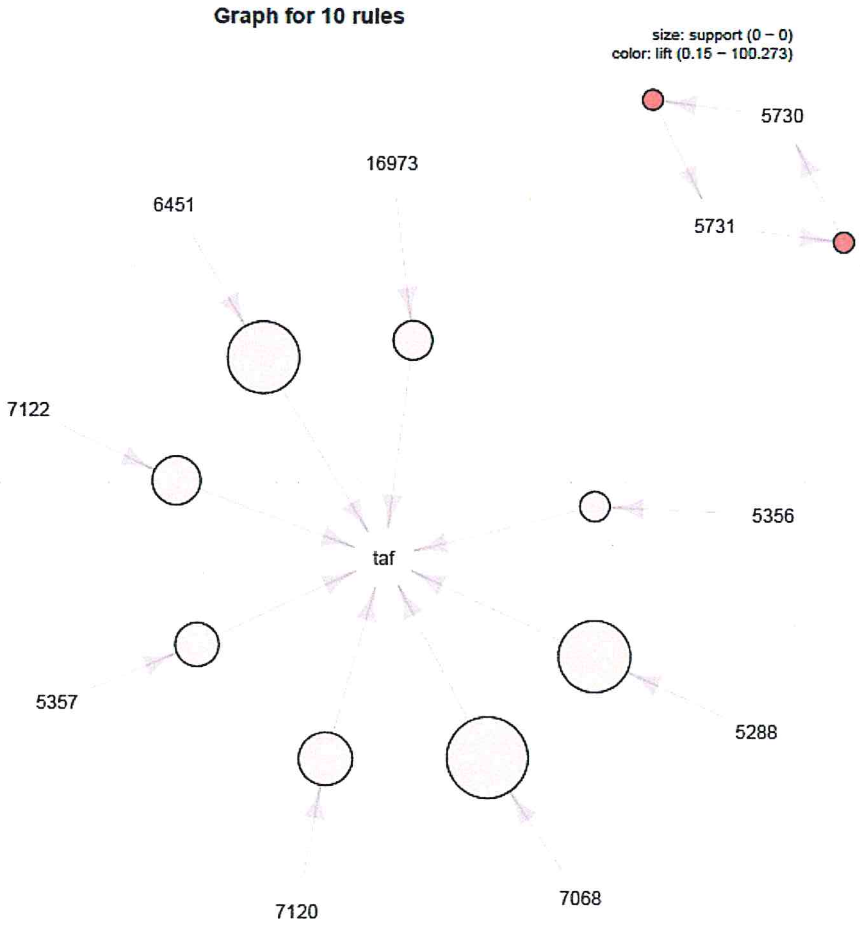


Figure 23: Best 10 Rules for Model No 4

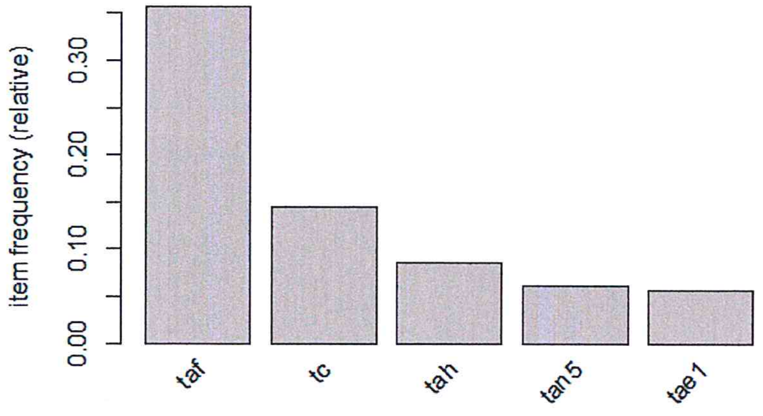


Figure 24: Most Frequent Items for Model No 4

APPENDIX E: RESULTS OF MODEL NO 5

The best 10 rules of model no 5 methodology are shown in Figure 25. Figure 26 is show us most 5 frequent items for the algorithm.

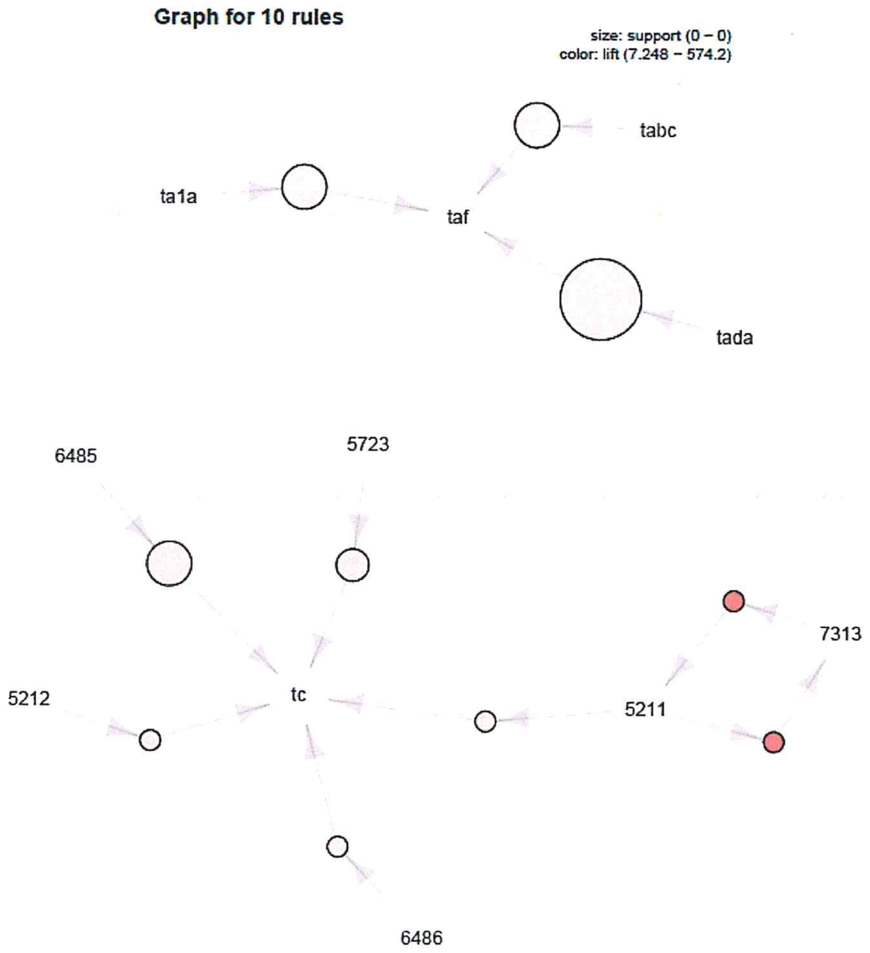


Figure 25: Best 10 Rules for Model No 5

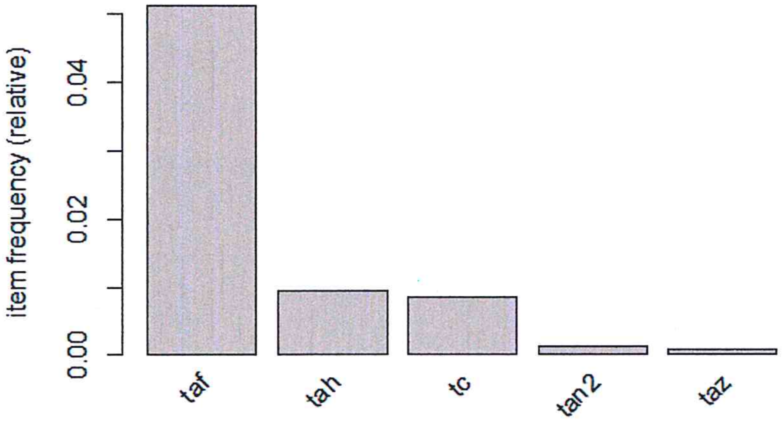


Figure 26: Most Frequent Items for Model No 5

APPENDIX F: RESULTS OF MODEL NO 6

The best 10 rules of model no 6 methodology are shown in Figure 27. Figure 28 is show us most 5 frequent items for the algorithm.

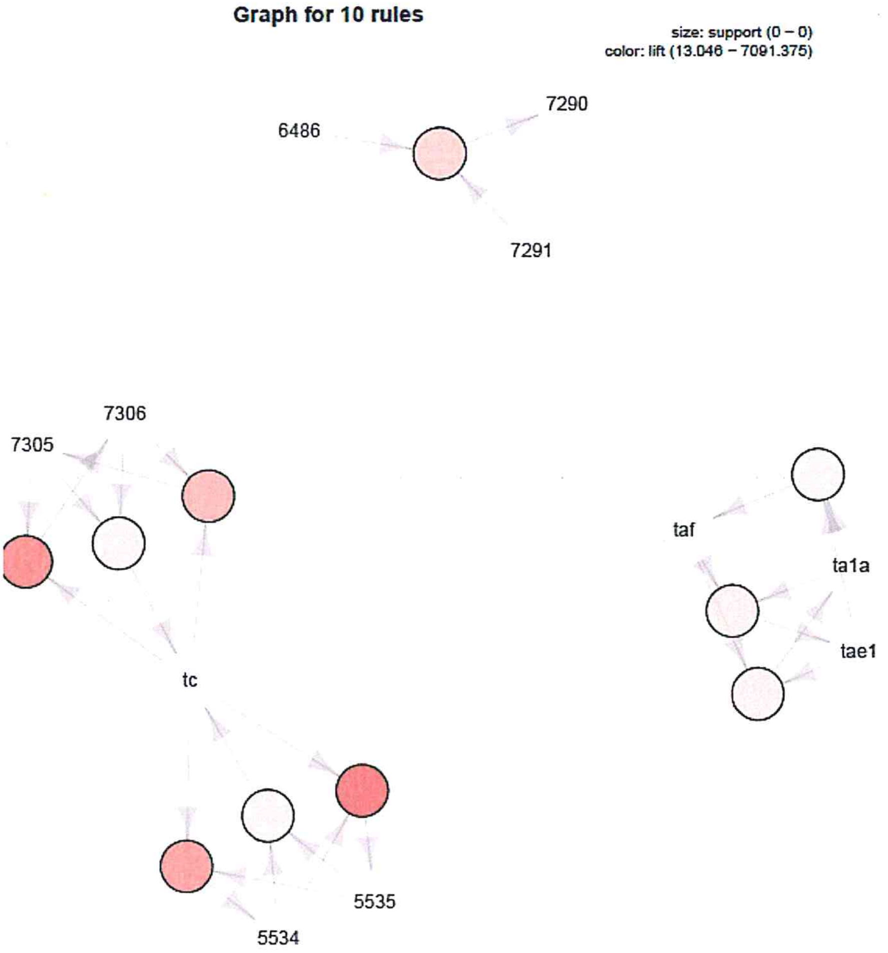


Figure 27: Best 10 Rules for Model No 6

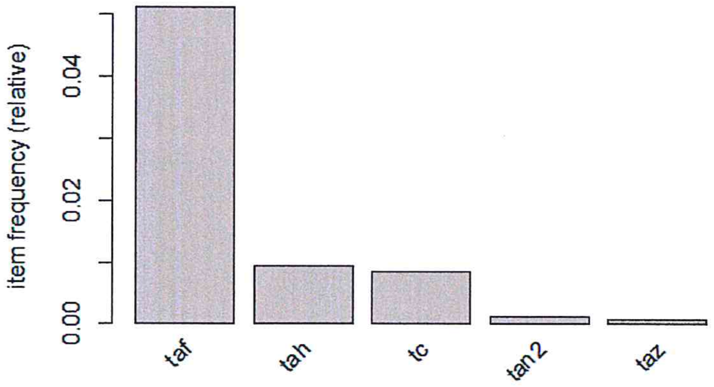


Figure 28: Most Frequent Items for Model No 6

REFERENCES

1- Berry, M.J.A., Linoff, G.S. (2004). Data Mining Techniques: for Marketing, Sales and Customer Relationship Management (second edition), Hungry Minds Inc.

2- Zaman, Ashrafi, T. David, K. Smith, ODAM (2004). An Optimized Distributed Association Rule Mining Algorithm, In IEEE Distributed Systems Online, Los Alamitos.

3- Yasemin Boztuğ, Thomas Reutterer. (2006). A Combined Approach for Segment- Specific Analysis of Market Basket Data.

4- Tan, Pang-Ning., Steinbach, Michael., Kumar, Vipin. (2006). Introduction to Data Mining, ch06, 327.

5- Ngai EWT, Xiu Li, Chau DCK. (2009). Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification Elsevier-Expert Systems with Applications

6- Katrin Dippold, Harald Hruschka. (2010). Variable Selection for Market Basket Analysis. University of Regensburg Working Papers in Business, Economics and Management Information Systems.

7- Charlet Annie M.C, Loraine., Kumar, Ashok., (2012). Market Basket Analysis for a Supermarket based on Frequent Item set Mining. ISSN(online): 1694-0814. pp 257

8- Verma Dipti, Nashine Rakesh. (2012). Data Mining: Next Generation Challenges and Future Directions - International Journal of Modeling and Optimization. pp 603

9- Intel. (2014). Getting Started with Big Data Analytics in Retail. Available From URL:https://www.intel.com/content/dam/www/public/us/en/documents/solution-briefs/retail-big-data-analytics-solution-blueprint.pdf

10- Lift in an association rule. (n.d.). Retrieved 07 23, 2013, from IBM. Available from URL: https://www.ibm.com/support/knowledgecenter/SSEPGG_9.7.0/com.ibm.im.model.doc/c_associations.html