A CASE STUDY: ANALYZE THE CUSTOMER RELATIONSHIP OF TNEB USING DATA MINING TECHNIQUES

B. Nevetha, B. Jasmine and M. Mareeshwari
Post Graduate Department of Computer Science
Jayaraj Annapackiam College for Women (Autonomous),
Periyakulum, Theni Dist, Tamil Nadu, India
jeromia2585@gmail.com

Abstract

Data mining techniques for automatic detection of relevant patterns in databases helps, businesses sift through layers of seemingly unrelated data for meaningful relationships. For the best results the mining process must be integrated with commercial data warehouses, and it must present the models in a relevant way for business users. The recent trends in the domain of power distribution and supply has been for State Electricity Boards to be split up into 3 entities – for generation, transmission and distribution. Tamil Nadu Electricity Board's (TNEB) total loss on the year 2013-2014 was 13,985.03 Crores. It is an increase of 20% when compared to the previous year. TNEB's debt, meanwhile, has gone up to 74,133.11 crores, which is more than 50% of the state's debt burden. These startling figures have come out in the latest (2013-14) revenue account and balance sheet of the power utility, which was uploaded on the TNEB website. There is a denunciation on the TNEB, which maintain poor customer strategy. This paper prepare proper customer relationship management strategies and implement them in full swing.

Keywords: *Datamining techniques, TNEB, Apriori Algorithm, Association rule mining.*

1. Introduction

The process, which an organization creates value for the customer is often referred to as Customer Relationship Management (CRM). According to Microsoft, CRM is a customer focused business strategy designed to optimize revenue, profitability, and customer loyalty. The economic environment on the last year

has dramatically modified, and the way in which companies interact with their customers has changed also. Because the actual conditions of customer stability is no longer guaranteed. Consequently, a company must understand their customers as better, and must respond quickly to their wants and needs.

Over the last few years, Tamil Nadu has been facing massive power deficits. According to the Central Electricity Authority (CEA), the state was expected to have a power deficit of around 18% in the year of 2010-11. As a result of this, the state is now facing huge power cuts. Averages of 3 to 4 hours of power cut are being experienced by the consumers in the state. Mainly the impact of this power shortage is being felt by the industries.

Data mining techniques for automatic detection of relevant patterns in database, helps businesses sift through layers of seemingly unrelated data for meaningful relationships. Data mining uses statistical and machine learning techniques to build models that predict customer behavior. For the best results the mining process must be integrated with commercial data warehouses, and it must present the models in a relevant way for business users.

2. Research and Methodology

The present study is an exploratory research conducted in the TNEB about their customer management. In order to pursue the aims and objectives outlined in the introduction, the content analysis of information gained from the multimedia research process was conducted to establish the underlying trends in location procedures used by the researcher.

The first stage is to gathering the secondary information from TNEB customers and formulating a location map of the selected brands of organized outlets. The second stage is to identifying the potential problems faced by the customers while approaching the EB offices for payment of bills and so on. They need to be categorized and organized in the flow of information.

The summary of interpretations was also given. In third stage, the profitability analysis was carried out by making specific assumptions in a hypothetical situation. In the last and the fourth stage, on the basis of the results

nd interpretations, specific postulates were framed, and on each postulate ypotheses were framed that can be tested through quantitative research in uture. The above-mentioned stages have been described as objectives in the receding paragraph.

.1. Apriori Algorithm

Apriori algorithm is a fundamental algorithm for mining association rule. t contains two processes:

- Detect all frequent item sets by scanning data base.
- Form strong association rules in the frequent item sets. Process one needs to scan DB several times, which consumes a lot of time and space. As a result, what needs to be improved is the mining competency of frequent group of things in DB.

.1.1. Key Concepts

Frequent Itemsets: The item sets which has minimum help (denoted by li or i^{th}-itemsets), Apriori property: any subgroup of frequent things must be requent.

Join Operation: to detect lk, a group of candidate k- group of things is leveloped by adding lk-1with itself.

2.1.2 Association Rule

Association rule of data mining involves picking out the unknown inter-dependence of the data and finding out the rules between those items [3]. Agrawal introduced association rules for point of sale (POS) systems in supermarkets. A rule is defined as an implication of the form A=>B, where A∩ B≠Ø. The left-hand side of the rule is called as antecedent. The right-hand side of the rule is called as consequent.

Support: $I = \{ i_1, i_2, i_3, \ldots , i_m\}$ is a collection of items. T be a collection of transactions associated with the items. Every transaction has an identifier TID [6]. Association rule A=>B is such that A Є I, B Є I.

A is called as Premise and B is called as Conclusion. The support ,S, is defined as the proportion of transactions in the data set which contains the itemset.

Support(X=>Y) = Support (XUY) = P(XUY).

Confidence: The confidence is defined as a conditional probability

Confidence (X=>Y) = Support (XUY) / Support(X) = P(Y/X). Lift: is the ratio of the probability that L and R occur together to the multiple of the two individual probabilities for L and R, i.e. lift = Pr(L,R) / Pr(L).Pr(R).

Conviction: is similar to lift, but it measures the effect of the right-hand-side not being true. It also inverts the ratio. So, a conviction is measured as:

Conviction = Pr(L).Pr(not R) / Pr(L,R)

2.1.3. Sample set

Association Rule Mining Algorithms in R:

APRIORI I a level-wise, breadth-first algorithm which counts transactions to find frequent itemsets and then derive association rules from them I apriori() in package arules.

```
>patterns = random.patterns(nItems = 150);
>summary(patterns);
>trans = random.transactions(nItems = 150, nTrans = 150, method = "agrawal", patterns = patterns);
>image(trans);
>tr<- read.transactions("Data1.csv", format = "basket", sep=',', rm.duplicates=TRUE)
>inspect(rules)
```

	lhs		rhs	support	confidence	lift
1	{}	=>	{0}	0.9880240	0.9880240	1.0000000
2	{}	=>	{1}	1.0000000	1.0000000	1.0000000
3	{15000}	=>	{0}	0.1556886	1.0000000	1.0121212
4	{15000}	=>	{1}	0.1556886	1.0000000	1.0000000
5	{10000}	=>	{0}	0.2934132	0.9865772	0.9985357

Data mining in TNEB

following are the benefits of adopting CRM processes in TNEB:

- **Develop better communication channels:** If we apply CRM in TNEB, it will help the department to develop the better communication channels in both ends.
- **Collect customer related data:** The TNEB has collected customer related data.
- **Create detailed profiles of individual customers:** There are maintain individual customer detail profile.
- **Increased customer satisfaction:** CRM has maintained customer satisfaction day by day increased.
- **Improved customer service and support:** TNEB also increased customer service.
- **Improved response time to customer requests for information:** If we request to management at any time.
- **Improved ability to meet customer requirements:** It can improve to meet and share the information.
- **Improved quality communication and network:** There is improved quality communicate.

.2.1. Data sources and methodology

arget population

This survey covers all the TNEB customers in Theni district.

nstrument design

This questionnaire collects data on experiences of customers when they approach any TNEB office irrespective of the geographical region. The items on the questionnaire have remained unchanged for several years. However, should modifications become necessary, proposed changes would go through a review committee, and a field test with respondents and data users to ensure its relevancy.

Sampling

This survey is a census with a cross-sectional design. Data are collected for all units of the target population, therefore no sampling is done.

Data sources

Responding to this, survey is mandatory. Data are collected directly from survey respondents. Data are compiled from the questionnaire collected. Respondents are sent a questionnaire to obtain their personal experiences on TNEB office. The researcher perform the data capture activities, and follow-up of non-respondents. Contact with respondents is maintained for subsequent follow-up.

Error detection

There are edits built into the data capture application to check the entered data for unusual values, as well as to check for logical inconsistencies. Whenever an edit fails, the interviewer is prompted to correct the information (with the help of the respondent when necessary). For most edit failures the interviewer has the ability to override the edit failure if necessary.

Imputation

A 100% response rate is attained; therefore imputation is not necessary.

2.2.2. Quality evaluation

Prior to the data release, combined survey results are analyzed for comparability; in general, this includes a detailed review of individual responses, general economic conditions, and historical trends. The data is examined at a macro level to ensure that the long-term trends make sense when compared to publicly available information in media reports, company press releases, etc. Subject matter officers follow up with the company to confirm the data and to document reasons for large fluctuations in sales or inventories.

2.2.3. Disclosure control

Releasing any data would divulge information obtained under the Statistics Act that relates to any identifiable person, business or organization

thout the prior knowledge or the consent in writing of that person, business or ganization. Various confidentiality rules are applied to all data that are leased or published to prevent the publication or disclosure of any information emed confidential. If necessary, data are suppressed to prevent direct or sidual disclosure of identifiable data.

2.4. Revisions and seasonal adjustment

Revisions in the raw data are required to correct known non-sampling rrors. These normally include replacing imputed data with reported data, orrections to previously reported data, and estimates for new births that were ot known at the time of the original estimates. Raw data are revised, on a nonthly basis, for the month immediately prior to the current reference month eing published.

The purpose is to correct any significant problems that have been found hat apply for an extended period. The actual period of revision depends on the nature of the problem identified, but rarely exceeds three years.

3. Analysis of Data

The researcher collects more than 400 responses from samples all over Theni district. She spent nearly 2 months to collect the fresh data from end users. After collecting the information, all the details are fed into the software and checked for outlier. The cleaned data was analyzed using single attribute and multiple attributes. Monthly Income, Electricity consumption and geographical location are considered as single attributes of the study.

4. Findings, Interpretations, Recommendations and Suggestions

4.1. Findings and Interpretations

The gathering of relevant and up-to-date information is a key business process. Information consists of organized facts and figures that have meaning within the context that the information is intended to be interpreted by people. Information is thus a valuable business commodity, and frequently businesses pay money for up-to-date and relevant information.

Consumption of Electricity: According to the district statistic book of Theni for the year 2012-13,

The power consumption of Theni district is 696.199 M.U

- Industries use143.041 M.U,
- Requirement for Agriculture & Huts is 293.688 M.U.,
- Domestic usages are 173.630 M.U.,
- Commercial usages of power is 43.840 M.U. and the Public lighting & Public Works consumes only 42 M.U.

The researcher found the power consumption from her survey also given below.

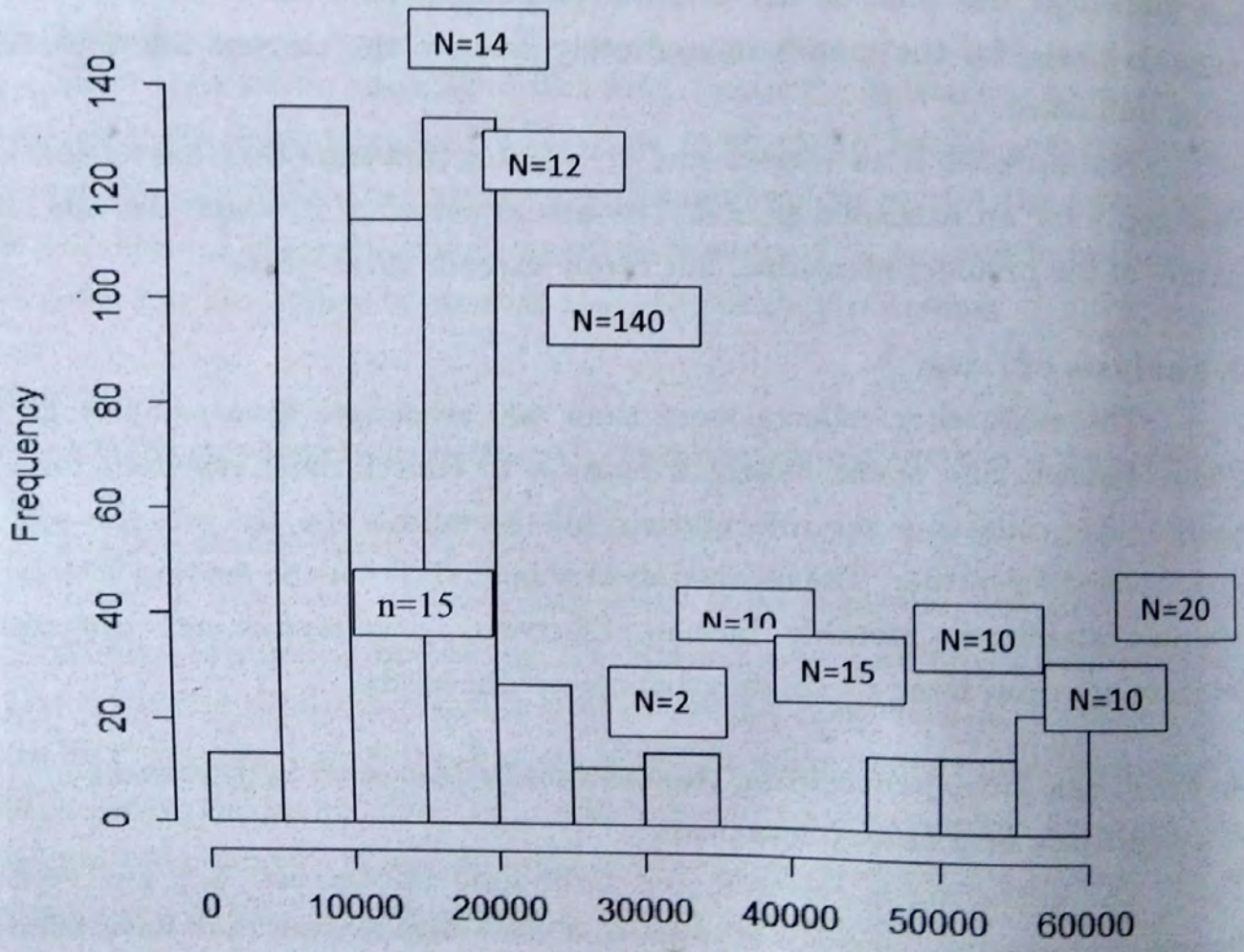


Fig 1. Survey of Power Consumption

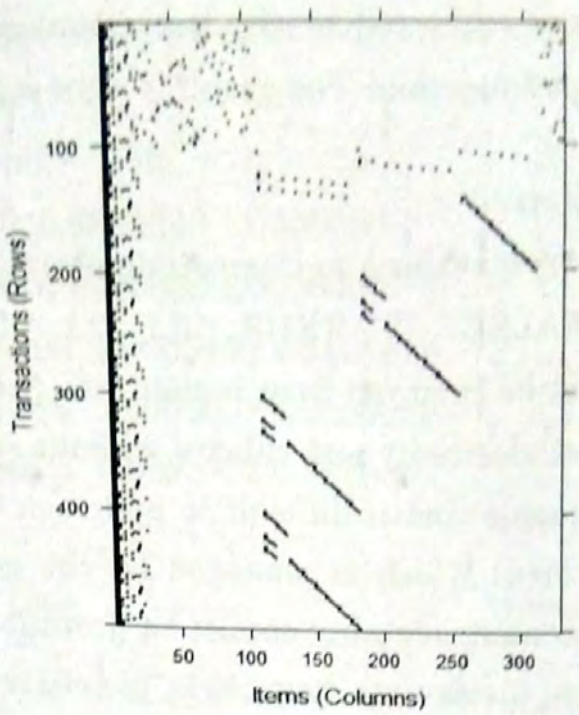


Fig 2. Graph for rural people power consumption

on Location: As the rural population of Theni district is more, the rural consumes more power and have more communication with the Electricity

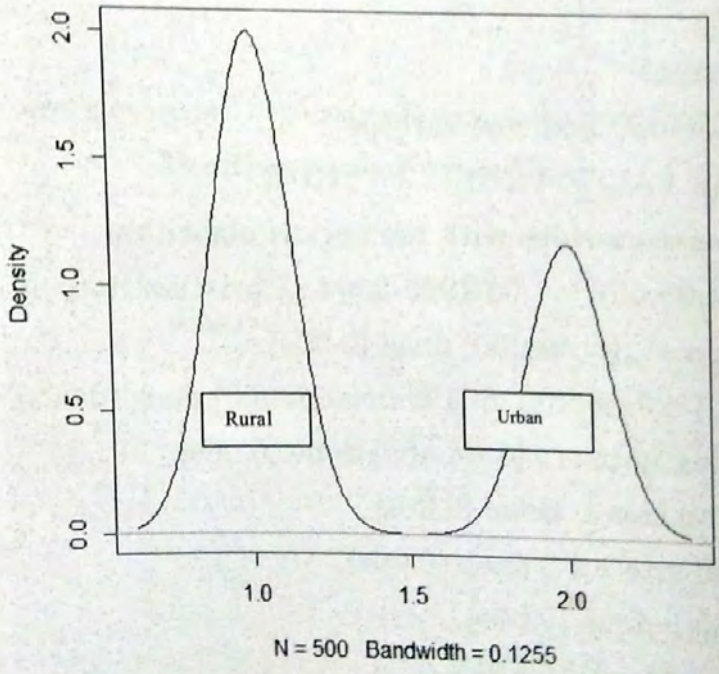


Fig 3. Graph for questionnaire collected from people

All the information collected by the questionnaire is represented in the form of graph and grouped together. The graph is represented as follows:

Parameter specification

Confidenceminvalsmaxaremavaloriginal Supportsupportminlenmaxlen target

0.8 0.1 1 none FALSE TRUE 0.1 1 10 rules

Electricity cannot be removed from human life. Still in Theni district, we can find villages without electricity and villages without street light. We can find poor illiterate village people find it difficult to approach the EB office for their requirement. A department which is managed by the government shows this much of non-customer center mentality cannot be profitable. This information is presented in line graphs, histograms, bar charts, pie charts etc. so that it is easy to quickly make sense of the information. Information processing refers to the collection, storage, manipulation, distribution and presentation of data, usually by electronic means. Computers are used widely in business for information processing. So the department can propose proper customer management strategies for its better revenue.

Algorithmic control

filter tree heap memopt load sort verbose

0.1 TRUE TRUE FALSE TRUE 2 TRUE

apriori - find association rules with the apriori algorithm

version 4.21 (2004.05.09) (c) 1996-2004 Christian Borgelt

set item appearances ...[0 item(s)] done [0.00s].

set transactions ...[329 item(s), 501 transaction(s)] done [0.00s].

sorting and recoding items ... [4 item(s)] done [0.00s].

creating transaction tree ... done [0.00s].

checking subsets of size 1 2 3 done [0.00s].

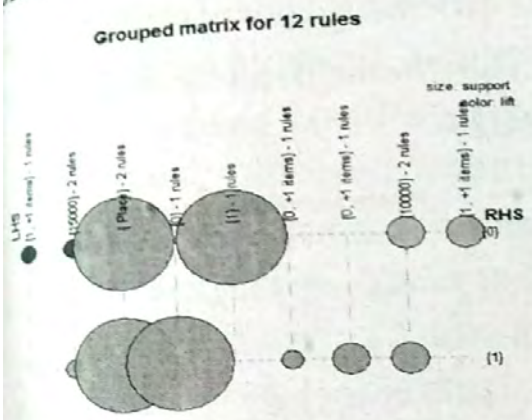
writing ... [12 rule(s)] done [0.00s].

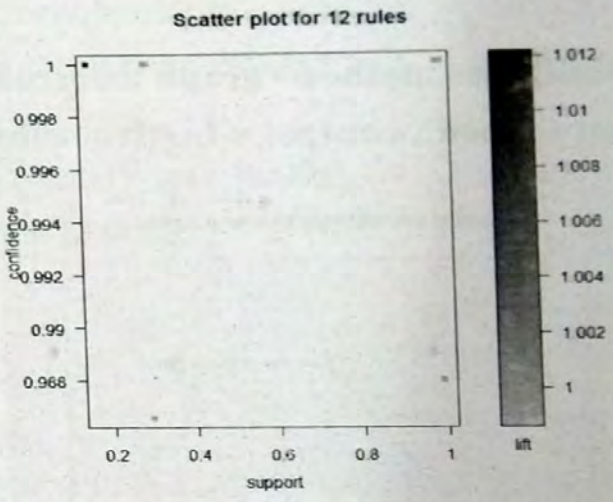
creating S4 object ... done [0.00s]

e support confidence level of data was analyzed for the EB data using
e statistical tool

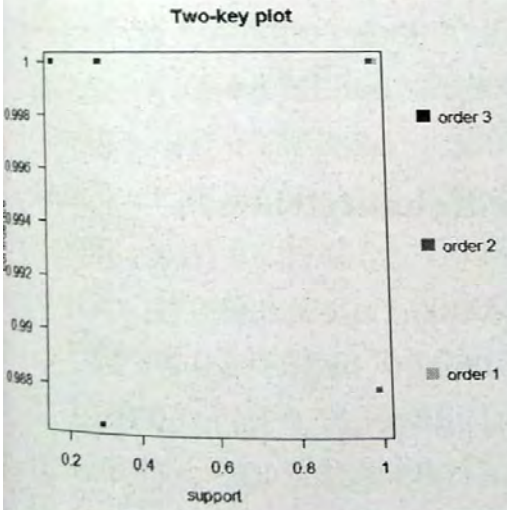
```
rhs  support confidence     lift
    => {0} 0.9880240  0.9880240 1.0000000
    => {1} 1.0000000  1.0000000 1.0000000
{15000} => {0} 0.1556886  1.0000000 1.0121212
{15000} => {1} 1.0000000  1.0000000 1.0000000
{10000} => {0} 0.2934132  0.9865772 0.9985357
```

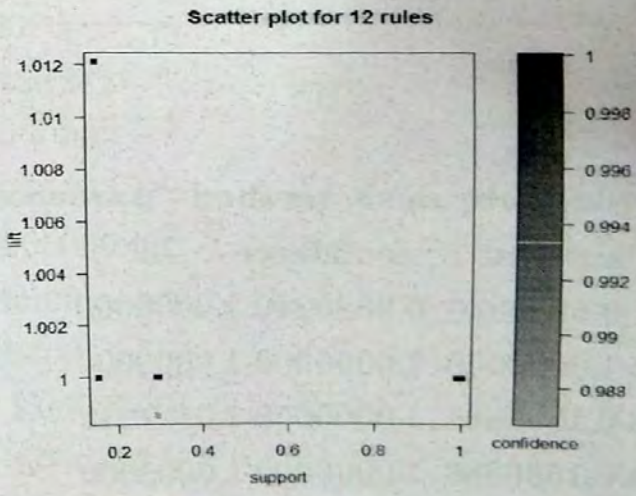
lot(rules)



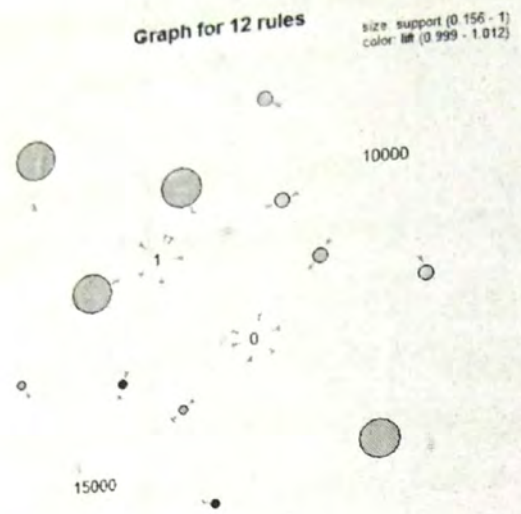


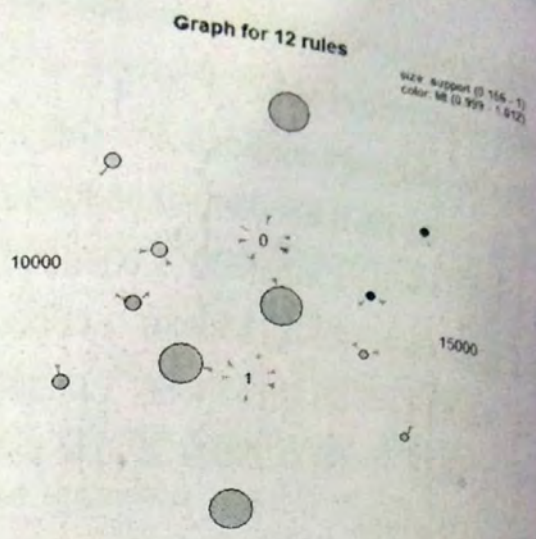
plot(rules,method="grouped")>plot(rules,measure=c("support","lift"), shading="confidence");



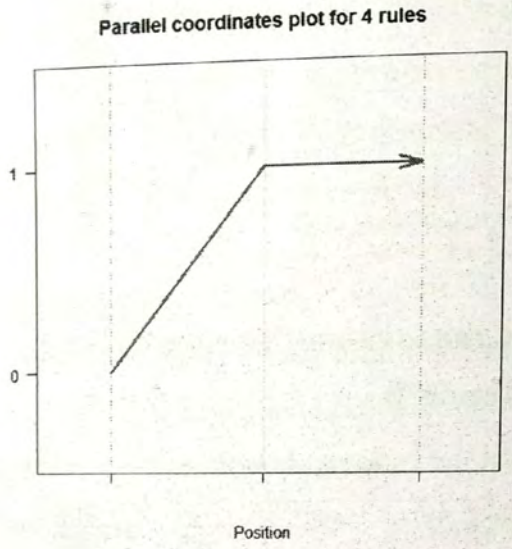


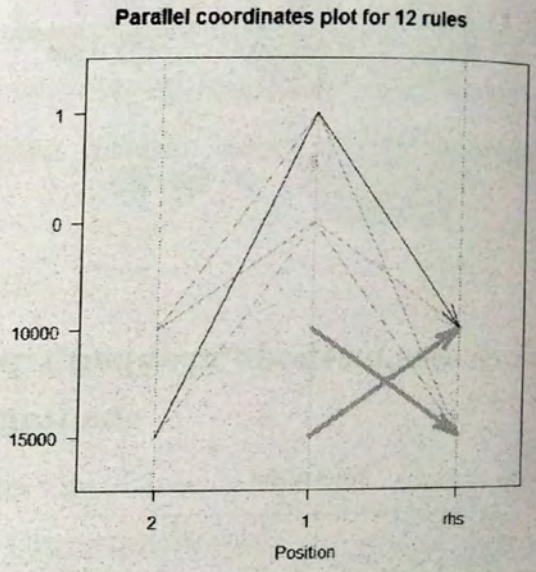
>plot(rules,shading="order",
control=list(main ="Two-key plot"));>plot(rules, method = "graph")





>plot(rules,method="graph",control=list(type="items"))>plot(rules,method= "paracoord", control = List(reorder = TRUE))





>plot(subrules2, method="paracoord");>head(quality(rules));

	support	confidence	lift
1	0.9880240	0.9880240	1.0000000
2	1.0000000	1.0000000	1.0000000
3	0.1556886	1.0000000	1.0121212
4	0.1556886	1.0000000	1.0000000
5	0.2934132	0.9865772	0.9985357

```
>inspect(oneRule);
lhsrhs support confidence lift
1 {} => {0}  0.99      0.99    1
>head(singleSupport, n = 5);
 576  955  739  317  185
0.122 0.084 0.074 0.062 0.063
```

Algorithmic control

```
filter tree heap memopt load sort verbose
   0.1 TRUE TRUE  FALSE TRUE    2    TRUE
apriori - find association rules with the apriori algorithm
version 4.21 (2004.05.09)        (c) 1996-2004   Christian Borgelt
set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[329 item(s), 501 transaction(s)] done [0.00s].
sorting and recoding items ... [4 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 done [0.00s].
writing ... [12 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s]
>inspect(rules)
lhsrhs   support confidence      lift
1 {}      => {0} 0.9880240  0.9880240 1.0000000
2 {}      => {1} 1.0000000  1.0000000 1.0000000
3 {15000} => {0} 0.1556886  1.0000000 1.0121212
4 {15000} => {1} 0.1556886  1.0000000 1.0000000
5 {10000} => {0} 0.2934132  0.9865772 0.9985357
6 {10000} => {1} 0.2974052  1.0000000 1.0000000
7 {0}     => {1} 0.9880240  1.0000000 1.0000000
8 {1}     => {0} 0.9880240  0.9880240 1.0000000
9 {0,  15000} => {1} 0.1556886  1.0000000 1.0000000
10 {1, 15000} => {0} 0.1556886  1.0000000 1.0121212
11 {0, 10000} => {1} 0.2934132  1.0000000 1.0000000
12 {1, 10000} => {0} 0.2934132  0.9865772 0.9985357
```

4.3. Conclusion and Future Analysis

In this present study, the effects of consumer reactions, value perception, trust perception and satisfaction on EB service were examined. The investigation was conducted in the population of Theni district, southern district of Tamilnadu. However, this investigation is suggested to be carried out in different cities and regions where geographical and demographical variables are considered to affect the variables that determine the loyalty of EB and CRM practiced everywhere. Therefore in this study on customer relationship in EB office, the variables of loyalty of EB officers can show differences. In our study, some effective variables about store loyalty or patronage behavior such as, social expectations, the social responsibility of the EB department, the cultural structure of customers were not included. For further studies, it is suggested that these variables be taken into consideration.

5. Reference

[1] Savitha S Kadiyala, Alok Srivastava, "Data Mining Techniques for Customer Relationship Management", IJTEL, ISSN: 2319-2135, VOL.6, December 2013.

[2] Berry, M., Linoff, G., (1997) Data Mining Techniques for Marketing, Sales and Customer Support, John Whiley & Sons.

[3] Alex Berson, Stephen Smith, Kurt Thearling, Building Data Mining Applications for CRMTMH Publications-2005.

[4] Jian Wang, Hua Ming Lu, Application of data mining technology in CRM, Journal of Beijing Information Science & Technology University, Vol 25,pp84—88, June 2010.

[5] Ruo Wu Zhong, Hui ping Wang, Research on CRM system based on data minin, Journal of Shaoguan University, Vol31, pp29—32, Mar 2010.

[6] Chowdhury Mofizur Rahman, Ferdous Ahmed Sohel, Parvez Naushad, Kamruzzaman S. M, "Text Classification Using the Concept of Association Rule of Data Mining," In Proceedings of International Conference on Information Technology, Kathmandu, Nepal, pp Arun K Pujari "Data Mining Techniques", Edition 2001.