DOI 10.4010/2016.1328
ISSN 2321 3361 © 2016 IJESC

Research Article Volume 6 Issue No. 5

Pattern Mining and Prediction in Shopping Mall

Ranjitha N[1], D.R.Umesh[2]
PG Scholar[1], Assistant professor[2]
Department of CSE
P.E.S.C.E, Mandya, India
ranjithan018@gmail.com[1], umesh.dr.pesce@gmail.com[2]

Abstract:
Real time shopping marketers are constantly looking for ways to improve the effectiveness of their campaigns. One way to do this is to target customers with the particular offers most likely to attract them back to the store and to spend money on their next visit. In this paper we divide the larger market into groups based on several defined criteria like age, gender, education etc. This information is stored in different datasets of database. This paper explains the different case study in order to explain theory of segmentation applied on datasets. The purpose of this case study is to determine dependency on products and shopping habits. Furthermore forecast sales determine the promotions of products and customer profiles. By observing customer buying patterns we can make the decision on product sales. Association rule mining was used as method for identifying customer buying pattern ans as a result customer profiles were determined. In this project RFID card reader is used to read the card number of customer, card ID is used for customer transaction details. At the same time customer receives message from connected mobile device. So that the customer will get the buying pattern which is based on previous shopping.

Keywords: Association rule mining, customer segmentation, Apriori algorithm, RFID

INTRODUCTION

Online shopping has become a trend nowadays rather than going out and buying items for themselves, reason being, online shopping provides an easier and quicker way to buy items and transactions are also quick when done online. The business entities also find easy to process the transaction based on the number of items sold to the users (i.e., bought by the users) through online transaction because all billing and other details are completely computerized which are saved in the database. Again coming to real time shopping, the entire transaction that happens in the mall is computerized and hence billing becomes easy. But in both cases, there arises an issue where the business entity needs to know the items that are being sold frequently. Not only the items frequently sold, but other items bought with that frequently bought item (Customer buying behavior) and also the moving behavior of the user.

This information is required in order to deal with the imports of items by the business entity depending on the sales record as well as placements of items in the Shopping Malls which increases sales. Now, how does one relate the items frequently bought with the arrangements/placements of items in Shopping Malls (Super Markets)? For which the system find the solution using a Data Mining Technique called Association Rule (pattern discovery).Besides association rules, interesting results were found about customer profiles, such as "What items do female customers buy?" or "What do consumers (married and 35-45 aged) prefer mostly?" . For instance, female customers purchase kurtha with a percentage of 60% whereas male customers purchase jeans with a percentage of 46%. Regarding to customers age, 65 and older customers purchase coffee with a percentage of 58%, and customers aged between 18- 25 preferred pizza with a percentage of 57%.The main goals of shopping is for Maximum profit, Minimum investment, Customer satisfaction.

II. STUDY

This part shows all steps from start to end of data mining process, including gathering raw data, normalizing, preparing raw data to be processed, as well as processing data with data mining software. After data is completely processed, results are shown in specific sub-part.

A. Data Preparation

Transaction data[1] is needed to produce association rules, which will be used to find frequent items that are purchase together. These data will be provided manually. Data is extracted from the data pool, filtered from several stores and customers. Raw data consist of 2 parts, one is customer information which has age, gender, and occupation, card no and other one is transaction data. Below are some examples for customer information and transaction data.

TABLE I. A PORTION OF CUSTOMER INFORMATION RAW DATA

CardNumber	Name	Gender	DateOfBirth	Address	ContactNo	EmailId	Occupation	Country
0001012274	Ramya	Female	1991	#432 North 2nd ...	+919738510586	ramya@gmail.com	Lecturer	India
0001797929	Rakesh	Male	1990	#59 1st stage vi...	+919986454547	rakesh@gmail.com	IT	India
1190988	Anil	Male	1995	#42,4th Block, V...	+919986454541	Anil@gmail.com	Student	India
1199001	Ajay	Male	1995	#122,4th Cross,...	+919978787445	Ajay@gmail.com	IT	India
1199005	Adithya	Male	1991	#22,7th Cross,r...	+919741201826	Adithya@gmail.c...	Doctor	India
1199007	Nandini	Female	1996	#1,5th Cross,ra...	+918197810320	Nandini@gmail.com	housewife	India
1199008	Chandini	Female	1985	#10,9th Cross,s...	+919986454547	Chandini@gmail....	Student	India
1199009	Ranjini	Female	1992	#12,5th Cross,e...	+919986398495	ranjini@gmail.com	Doctor	India
1199010	prem	Male	1990	#72,1th Cross,j...	+919448273367	prem@gmail.com	Student	India
1199012	Nisarga	Female	1994	#12,15th Cross,...	+918088533916	Nisarga@gmail.c...	Student	India
1199019	Rakshitha	Female	1988	#7,1st Cross,hr...	+918792235221	Rakshitha@gmai...	Bank	India
1199020	pooja	Female	1990	#2,5th Cross,br...	+917411918565	pooja@gmail.com	Student	India
1199024	Darshan	Female	1987	#1,15th Cross,n...	+919738510586	Darshan@gmail....	Student	India
1199026	Mounisha	Female	1997	#2,10th Cross,k...	+919900658853	Mounisha@gmail....	Student	India
1199027	krutik	Male	1986	#3,11th Cross,c...	+919739836975	krutik@gmail.com	Bank	India
1199039	Shashank	Male	1980	#15,3rd Cross,r...	+919181478727	Shashank@gmail....	Bank	India
1199061	Abhi	Female	1980	#35,25th Cross,...	+9194144677686	Abhi@gmail.com	IT	India
1199078	Ayesha	Female	1992	#101,22nd Cros...	+919754335787	Ayesha@gmail.c...	Student	India
1199079	Nischal	Male	1984	#102,23rd Cros...	+919755536888	Nischal@gmail.com	engineer	India
1199148	Amith	Male	1991	#22,7th Cross,r...	+919876543246	Amith@gmail.com	Mechanical engg	India

TABLE II. AN EXAMPLE OF CUSTOMER TRANSACTION RAW DATA

Details_ID	Transaction_ID	Item_ID	Quantity
4	3	5	1
5	3	9	1
6	3	11	1
7	4	5	1
8	4	4	1
9	4	8	1
10	5	4	1
11	5	5	1
12	6	6	1
13	6	1	1
14	7	1	1
15	7	2	1
16	7	6	1
17	6	3	1
18	6	16	1
19	6	14	1
20	6	18	1
21	6	17	1
22	7	17	1
23	7	16	1
24	8	1	1
NULL	NULL	NULL	NULL

B .Preparation of Raw Data

Data mining softwares cannot process raw data, as they need tabular forms in order to mine. It is needed to convert these raw data to such a form so that softwares can process. Below are the examples of converted raw data, ready to process.

Item_ID	SubCategory_ID	Item_Name	Item_Cost	Item_Details	Item_Image	Quantity	Investment
1	1	Cotton kurthis	2121	Cotton kurthis	~/Item_Photos/...	1	0
2	1	long kurthis	1444	Cotton kurthis	~/Item_Photos/l...	1	0
3	36	3D Sound system	16799	Philips Sound sy...	~/Item_Photos/...	3	0
4	45	Carrot	43	ooty carrot	~/Item_Photos/...	3	0
5	46	Apples	124	KAshmir apples	~/Item_Photos/...	8	0
6	1	Designer kurthis	2134	DEsigner kurthis	~/Item_Photos/...	2	0
7	47	Lays	56	Lays	~/Item_Photos/...	5	0
8	47	Choclates	34	Choclates	~/Item_Photos/...	2	0
9	45	Onion	45	Onion	~/Item_Photos/...	7	0
10	45	Cabbage	38	cabbage	~/Item_Photos/...	2	0
11	46	grapes	120	grapes	~/Item_Photos/...	5	0
12	46	oranges	145	oranges	~/Item_Photos/...	4	0
13	38	Jeggins	4567	Jeggins	~/Item_Photos/...	4	0
14	33	whirlpool refrige...	25788	whirlpool refrige...	~/Item_Photos/...	2	0
15	34	nokia 1100	1500	nokia 1100	~/Item_Photos/...	6	0
16	36	Philips home the...	10000	Philips home the...	~/Item_Photos/...	7	0
17	35	Moto G	12000	Moto G android ...	~/Item_Photos/...	7	0
18	35	HTC desire	20022	HTC desire 826	~/Item_Photos/...	4	0
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

III.METHAODLOGY

Association Rules

Apriori TID Algorithm

*STEP 1:*Scan the customer transaction data set and determine the support(s) of each item.

*STEP 2:*Generate L1 (Frequent one item set).

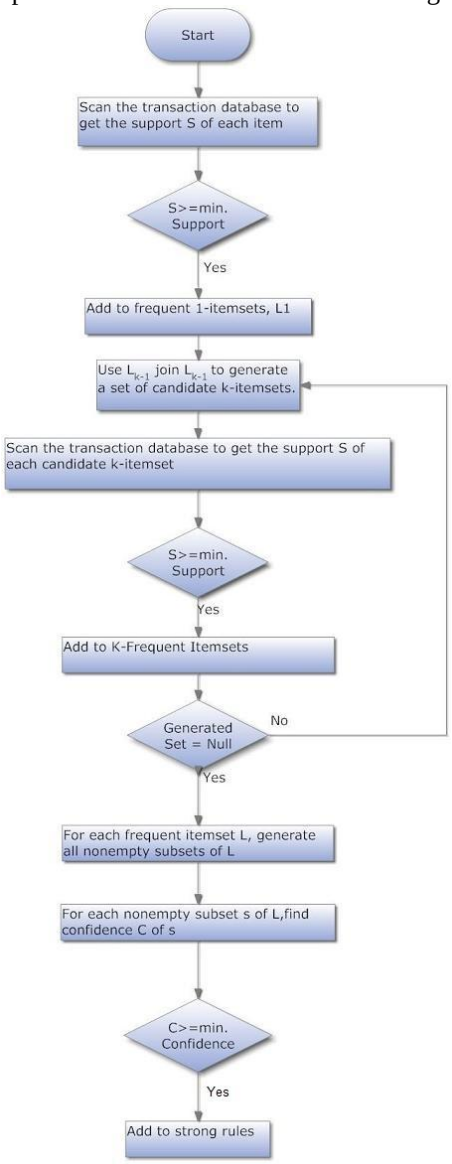
*STEP 3:*Use Lk-1, join Lk-1 to generate the set of candidate k - item set.

*STEP 4:*Scan the candidate k item set and generate the support of each candidate k – item set.

*STEP 5:*Add to frequent item set, until C=Null Set.

*STEP 6:*For each item in the frequent item set generate all non-empty subsets.

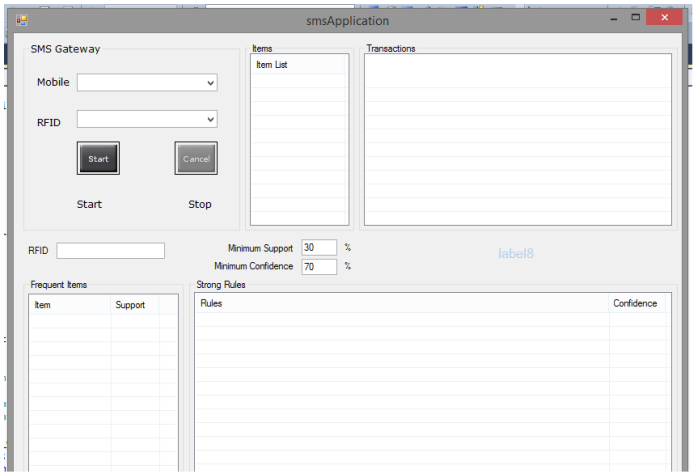
*STEP 7:*For each non empty subset determine the confidence. If confidence is greater than or equal to this specified confidence .Then add to Strong Association Rule.



IV. APPLICATION

The below screenshot shows the GUI of SMS application in this application we first select the port numbers of mobile and RFID reader and click on the start button which runs the application for continuously 24 hours. When the person

sense the RFID Card the ID number will be noted and the strong rules generated will be displayed here and sent as SMS for customer mobile. The RFID Card holds the unique ID and customer details like name, date of birth, gender, occupation, address etc.



V. RESULTS

With the advance of technology, databases are becoming more and more important for current information technology. Databases stores huge amount of data and data mining allows extracting valuable information from this datasets. There are many techniques to mine these data and association rule mining is one of the most important among these. Apriori algorithm[1] is one of the most important tools for association rule mining. In this study, Apriori algorithm is applied for mining association rules in database. This database included customer informations like age and gender but without confidential information like name, address and phone. A unique customer ID is used to distinct customers.

Performance Factor

- Data Structure – array based
- Memory Utilization – depends on the data set [less for small dataset]
- Number of scans - multiple scan to generate candidate set
- Execution time - execution time depends on producing candidates

VII. CONCLUSION

The system discovers patterns using the data mining technique called as “Association Rules”. The system aims at developing pattern mining and prediction techniques that explore the correlation between the moving behavior and purchasing transactions of mobile users to explore potential features. Discovering useful patterns hidden in a database using utility pattern mining. Mining high utility item sets from databases refer to finding the itemsets with high profits. Relative importance of each item and quantities of items should be considered. Finally this system customer buying patterns.

VII. FUTURE WORK

This paper is done for pattern mining and prediction based on date, gender and age. Here the pattern mining means customer purchasing behavior. In future work pattern mining can be done based on customer occupation, marital status etc.

VIII. REFERENCES

I. Sabri Serkan Güllüoğlu Computer Engineering Department Istanbul Arel University İstanbul, Turkey **“**Segmenting customers with data mining techniques”,2015

II. L. B. Cristopher, “Mining Rules In Single-Table and Multiple-Table Databases”, 2002.

III. R .Agrawal, R. Srikant, “Fast Algorithms for Mining Association Rules in Large Databases", Proceedings of the 20thInternational Conference on Very Large Data Bases (VLDB),Santiago, Chile, September 1994.

IV. G. Shaw, “Discovery & Effective Use Of Quality Association Rules in Multi-Level Datasets”, 2010.

V. H. Mannila, “Methods and Problems in Data Mining”, Proceedings of the 6th International Conference on Database Theory, 1997.

VI. M. Berry, G. Linoff, “Data Mining Techniques for Marketing, Sales, and Customer Relationship Management”, Second Edition, John Wiley & Sons, 2004.

VII. D. E. Sharp, “Customer Relationship Management Systems Handbook”, CRC Press, 2003.

VIII. G. Bukhbinder, M. Krumenaker, A. Phillips, “Insurance Industry Decision Support: Data Marts, OLAP and Predictive Analytics”, 2005.

IX. G. Galfond, “Data Mining Can Unearth A Competitive Edge”, October 6, 1997.

X. R. Wirth, J. Hipp, “CRISP-DM: Towards a Standard Process Model for Data Mining”, Proceedings of the Fourth International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 2000.

XI. D. T. Larose, “Discovering Knowledge in Data: An Introduction to Data Mining”, John Wiley & Sons, 2005.

XII. B. Moxon, “Defining Data Mining”, DBMS Data Warehouse Supplement, August, 1996.

XIII. R. Agrawal, A. Bollinger, T.Mehta, M. Shafer, J. and Srikant, “The Quest Data Mining System”, Proc. of the 2nd lnt'l Conference on Knowledge Discovery in Databases and Data Mining, Portland, Oregon, August 1996.

XIV. R. Agrawal, T. Imielinski, and A.Swami, “Database Mining: A Performance Perspective”, IEEE Transactions on Knowledge and Data Engineering, Special issue on Learning and Discovery in Knowledge-Based Databases, 5(6): 914-925, December 1993.

VI. ACKNOWLEDGEMENT

We would like to thanks our facility member for theirsupport and help for our research activities. And also to our family members for their moral and mental support