

Segregating Customers of Retail Market based on their Purchasing Pattern

Supriya V Koganurmath¹ Madhushree²

¹M.Tech. Student ²Assistant Professor

^{1,2}Department of Computer Science & Engineering

^{1,2}SCE College of Engineering, Bangalore

Abstract— Retail advertisers are continually searching for ways to enhance the adequacy of their crusades. One approach to do this is to target clients with the specific offers no doubt to pull in them back to the store and to invest more energy and cash on their following visit. Demographic business sector division is a way to deal with fragmenting markets. An organization partitions the bigger business sector into gatherings in view of a few characterized criteria. Age, sexual orientation, occupation are among the generally considered demographics division criteria. A specimen contextual analysis has been done with a specific end goal to clarify the hypothesis of division connected on a retail market chain. The reason for this contextual analysis is to decide reliance on items and shopping propensities. Besides estimate deals decide the advancements of items and client profiles. Association rule mining was utilized as a technique for distinguishing clients purchasing designs and as a result client profiles were resolved. Other than Association rule, fascinating results were found about client profiles, for example, "What things do female clients purchase?" or "What do consumers (married and 35-45 matured) lean toward for the most part?".

Key words: Data Mining, Association Rule Mining, FP-Growth

I. INTRODUCTION

A definitive objective of choice emotionally supportive networks is to furnish directors with data that is valuable for understanding different administrative parts of an issue what's more, to pick a best arrangement among numerous choices. In the work, it is concentrated on a certain choice emotionally supportive network in the interest of business sector chiefs who need to create and actualize effective promoting programs by completely using a client database. This is critical on the grounds that, because of the developing interest in miniaturized scale showcasing, numerous organizations commit significant assets to distinguishing family units that might be open to focused promoting messages. This turns out to be more basic through the simple accessibility of information distribution centers joining demographic, psychographic and behavioral data.

II. LITERATURE SURVEY

A. "Fast Algorithms for Mining Association Rules in Large Databases" [1]

The work considers deals exchange in which the issue of finding affiliation rules for things in an extensive database. In the work, the apriori calculation has been utilized to mine Association Rules. The calculation which makes different sweeps over database to discover the regular things and it produces an excessive number of competitor things which

are some of the time not solid for further utilization, similar to this the calculation continues creating the applicant things which are huge and calculation figures the viable backing for these thing sets. AIS calculation has been utilized as a part of the study to create Candidate itemsets, Candidate itemsets are produced and considered the database is checked.

1) *Limitations*

Sequential programming design have been followed, devours additional time.

B. "An Intelligent System for Customer Targeting"[2]

The artificial neural systems (ANNs) guided by genetic algorithm (Gas) have been utilized as a part of Data digging strategies for business sector supervisors. To pick the ideal arrangement the point of the emotionally supportive network is to extricate helpful essential data to know the managerial angles. By using the client information the work focuses on choice emotionally supportive network to man-age the viable battles on advertising programs.

1) *Limitations*

– Over fitting

C. "Customer Segmentation Using Clustering and Data Mining Techniques" [3]

It sorts crude information into important bunches and gatherings of moderately homogeneous perceptions. The objects of a specific group have comparative qualities and properties however vary with those of different bunches. The gathering is expert by discovering similitude among information as per qualities found in crude information. K-Means grouping Technique have been utilized, the primary goal was to discover ideal number of bunches. The central capacity of calculation involves finding the k-means. Initial, an underlying arrangement of means is characterized and after that resulting grouping depends on their separations to the centers.

1) *Limitations*

– Reliance on users to specify number of clusters.

Author Name	Algorithms and Methods	Drawbacks
R Agrawal, R Srikant	Apriori Hybrid, AIS	Sequential Programming, Consumes more time
YongSeon Kim	Genetic algorithm	Over fitting
Krishana R. Kashwan, C. M Velu	K-Means clustering Technique	Depends on the user to give the number of clusters at the early stage

Table 1: Comparisons of Existing Systems

III. PROPOSED SYSTEM

The Proposed system shows working of customer segregation

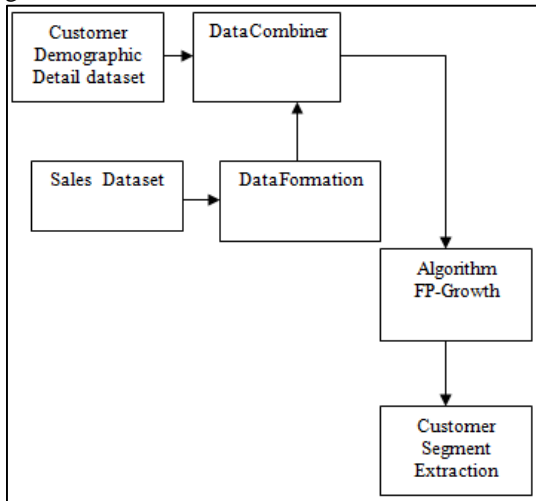


Fig. 1: System Architecture

There are 6 steps in the proposed system those are as follows:

A. Customer Demographic Detail Dataset

The customer demographic detail dataset contains customer information like age group, sex, profession. It collects information like this in the retail markets.

B. Sales Dataset

Sales dataset contains information of all the items purchased by the customer. It is nothing but the transaction details.

C. Data Formation

Data formation where horizontal data is converted to vertical format.

This module converts the transaction data in horizontal format to vertical format. Where 1 indicates customer has bought that product, 0 indicates customer has not bought that product.

Items purchased Data in horizontal format										
1	2	3	4	5						
2	4	5	6	7	8	9	10			
2	5	7								
3	4									
1	4	6	8	9						
8	9									

Each row represents items purchased.
The row number is considered as "items purchased list ID"

Table 2: Items purchased Data in horizontal format

Items purchased Data in Vertical Format

Item ID	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	0	0	0	0	0
2	0	1	1	1	1	1	1	1	1	1
3	0	1	0	0	1	0	1	0	0	0
4	0	0	1	1	0	0	0	0	0	0
5	1	0	0	1	0	1	0	1	1	0
6	0	0	0	0	0	0	0	1	1	0

Table 3: Items purchased Data in Vertical Format

D. Data Combiner

It combines both the customer and sales data information.

E. FP-Growth:

The combined data will be mined to get frequent patterns, after getting frequent patterns association rules will be generated.

1) Algorithm: FP-Growth

Input: Sales data set
Output: Customer Segregation
Step1: Get all the customer information data and purchased items data
Step2: Combine both the customer data and purchased items data
Step3: Apply FP-Growth algorithm. if (mapSupport.get(item)>=minSupportRelative), add items that satisfies minimum support.
Step4: Generating the association rules over the frequent items to find out the customer likings.
Step5: Finally segregating the customers based on their purchasing patterns.

Table 4: Proposed Algorithm

IV. RESULTS AND ANALYSIS

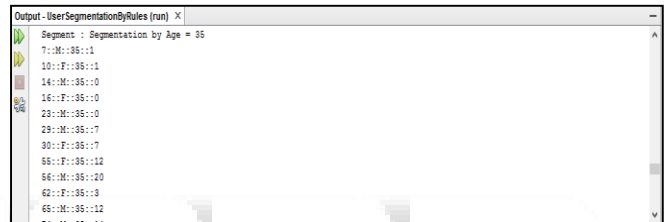


Fig. 2: Segregation by age



Fig. 3: Segregation by occupation

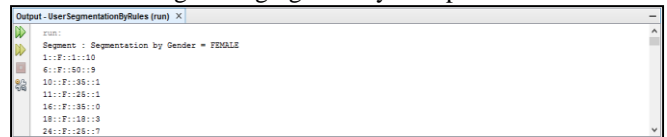


Fig. 4: Segregation by gender

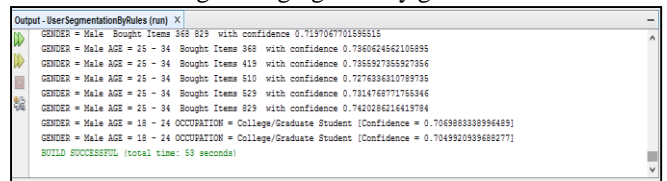


Fig. 5: Customer Segregation

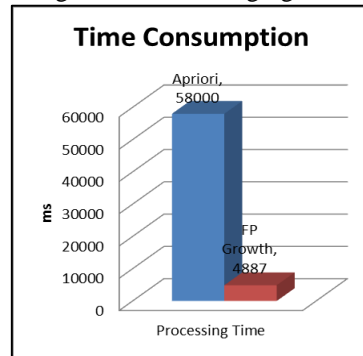


Fig. 6: Time Consumption

The result shows different groups based on customer data and also the product has been purchased regularly with confidence value, Result also shows comparison made between Apriori algorithm and FP-Growth algorithm, for 5000 transaction data the algorithm is calculated, it shows that FP-Growth algorithm is more effective and less time consuming than existing apriori algorithm.

V. CONCLUSION

With the assistance of these outcomes, much helpful data can be found, the calculation shows FP-Growth is more faster than other algorithms, it effectively scans entire database to find out frequent patterns. From the result it is helpful to conclude most sold things can be in advancement together to build deals by pulling in customers. Minimum sold things can likewise be joined with most sold things to expand offers of it. Besides, store designs can be improved to expand deals, shopping time of customers.

ACKNOWLEDGEMENT

I express deep thanks to Mrs. Madhushree, Assistant Professor, Dept. of CSE for her valuable advice and support provided through out for completing this paper. I thank the anonymous referees for their reviews that significantly improved the presentation of this paper. Words cannot express our gratitude for all those people who helped us directly or indirectly in our endeavor. I take this opportunity to express my sincere thanks to all staff members of CS&E department of SCE for the valuable suggestion.

REFERENCES

- [1] R.Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Santiago, Chile, September 1994.
- [2] YongSeon Kim, W. Nick Street "An Intelligent System for Customer Targeting" Business information system, Utah state University, Logan, USA, December 2002.
- [3] Krishana R. Kashwan, C. M Velu (2013) "Customer Segmentation Using Clustering and Data Mining Techniques" International Journal of Computer theory and Engineering, IACSIT, 2013.
- [4] Blattberg, R., Buesing, T., Peacock, P., & Sen, S. (1978). Identifying deal prone segment. *Journal of Marketing Research*, 15(3), 369–377.
- [5] Bult, J. R., & Wansbeek, T. (1995). Optimal selection for direct mail. *Marketing Science*, 14(4), 378.
- [6] Dibb, S., & Simkin, L. (1997). A program for implementing market segmentation *Journal of Business and Industrial Marketing*, 12(1), 51–65.
- [7] Frank, R. E., Massy, W. F., & Boyd, H. W. (1967). Correlates of grocery product consumption rates. *Journal of Marketing Research*, 4(2), 184–190.
- [8] Green, P. E., & Wind, Y. (1973). *Multiattribute decisions in marketing*. Dryden Press.
- [9] Haley, R. I. (1968). Benefit segmentation: A decision-oriented research tool. *Journal of Marketing*, 32(3), 30–35.
- [10] Helsen, K., & Green, P. E. (1991). A computational study of replicated clustering with an application to

market-segmentation. *Decision Sciences*, 22(5), 1124–1141.

- [11] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM*