

http://www.mapletreejournals.com/index.php/irijmtReceived 05 March 2020ISSN 2582-1040Accepted 24 May 20202020; 2(3);14-21Published online 30 May 2020

Customer Relationship and Sales Management using Data Mining Techniques

M. Devaganesh ^{1,*}, B. Dilipan ¹, R. Abirami ¹

¹ Department of Information Technology, Rajalakshmi Institute of Technology, Chennai-600124, Tamil Nadu, India. *Corresponding author E-Mail ID: <u>devaganeshritit@gmail.com</u>

DOI: https://doi.org/10.34256/irjmt2033

Abstract

A general store is a destination for customers to buy everything they require in a single spot. The Manager is responsible for administering and making decisions for developing and running the store. But it is a heavy task for a single person to make valuable decisions that do not affect the business. So, in this project, data mining techniques are applied to the store's transactional data to help the manager make decisions. We use RFM, Recency Frequency and Monetary segmentation method and classification algorithms to classify the customers into loyal and not-so-loyal customers to make predictions based on the data of the valuable customers. The customers who are loyal to the store are selected and their data is used for further analytics. Association mining technique to obtain the products which are most likely to be bought together and other data mining and visualization techniques to display the valuable knowledge mined from the dataset. Then we cluster the products based on the product's important keywords and then the customers are further classified by finding the number of products bought from the previously found clusters.

Keywords: Products, Classification, Segmentation, RFM, Clustering.

1. Introduction

Stores that would sell various kinds of products that the customer may require are called General Stores or Supermarkets. Due to the growing population and the increasing requirements of customers from General Stores, it has become tough for the general stores to attain the credibility of their customers. As the reliance on General Stores has become high for the people, it has become the general store's responsibility to know their customer's requirements and to attain them. For the successful management of the General Stores, it is needed to make the right decisions at the right times, which is invariably being done by the store manager, the accuracy of the decisions made is limited to the individual's ability and there is a high possibility of subjective errors. To reduce the risks due to the above-mentioned factors, it is required to automate the decision-making processes of General Stores using data mining algorithms.

1.1 Goal

Understanding our customers and making decisions based on their requirements will be enabling the manager to make appropriate decisions for the General Stores processing. This model intends to find out the customers who are more impactful in the sales of the stores and find out the insights about their purchase pattern. Making such refinement of the customer will reduce the gap between the actual occurrence and the perception of the sales in-store. Making the appropriate perception will lay a path for deciding on things that would be having a greater impact on the store's progression. Dataset consists of attributes of every purchase that are made in the store, which includes detail of the customer who made the purchase, the product details. Upon the dataset, the customers with high relevance called as "loyal" customers are found out with their purchase details mapped along. With the refined details, further segmentations are made to find out the insights.

1.2 Literature Survey

With the increasing number of competitions for a general store, the store manager must make valuable decisions for the store to attract new customers and to retain existing customers. But with a large number of customers, the task of making business decisions that is suitable for the customers as well as for the development of the business is almost impossible for a single manager. This paper tells the methods and the techniques which can be employed for easing the manager's workload.

[1] [2] The best way of doing this is to segment the customers based on their loyalty towards the store. [3] A business principle - '80/20 rule' states that 80% of a business's profit is from 20% of its customers [3]. So, it is enough to target the loyal customers for the developing prediction models for the development of the business instead of targeting every customer. [1] Also we have to understand that not all customers who spend much in a store are loyal customers and not all customers who spend less in a store are disloyal customers [1].

[4-7] We can segment the customers based on their loyalty into two segments - The Loyal Customer and The Not-So-Loyal Customer using the RFM Segmentation technique. RFM segmentation is a method where a customer is given a score based on how recently they visited the store - Recency, how frequently they visit the store - Frequency, and how they spend in the store - Monetary [4] [7]. Based on their RFM score the loyalty is decided by keeping a threshold loyalty score. Customers above the threshold are considered to be Loyal and the rest as Not-Loyal. Now the data of the Loyal customers are considered for the rest of the data mining process [3].

[8] Products sold in the store are initially clustered based on their properties. The Loyal customers are further classified based on the cluster from which they have bought the most number of products in that store [9]. So, the number of classifier labels depends on the number of product clusters formed previously.

[10] The loyal customers' data is again used to find the association or relationship between the products. [11] We can get information about the products which are often bought together by the customers. Using this data, the store manager can arrange these products on the shelves nearby which increases the probability of selling these products to a single customer.

1.3 RFM Classification

Customers of a General Store may differ drastically from one another, some may have a greater contribution to the store's sale and others may have a much lesser contribution. To make decisions based on customer buying patterns at first it is required to find whom to be considered and whom to be left behind, to do these we use a model called as R-F-M (Recency - Frequency - Monetary) model. Which would consider recency, frequency and monetary contribution of each customer to derive an RFM score for every customer, based on the RFM score the customers are classified into loyal (who are considered for decision making) and not-so-loyal (who are not considered for decision making). We generate a classification model to predict the loyalty of future customers.

1.4 Product Segmentation

For making decisions in General Stores, it is required to consider the product which is having a pattern in the customer's purchase history, those products can be considered to complement each other or belonging to a particular set of purchases. This process of clubbing products that are relevant to each other is called segmentation. The products are segmented so that decision making for a single product in the segment will work out in a similar way collectively for other products that are present in the same segment as the former.

1.5 Customer Classification

Now the loyal customers are once again segmented based on the products they buy. A customer might have bought many products. Every individual product that a customer bought will fall under any one of the product segments. The products under which segment the customer mostly bought are founded out and it is used to segment the customers. A classification model will be developed to segment the future customers.

1.6 Product Association

The loyal customers' purchase data is used to find the association or relationship between the products in the store. By doing so we can find the products which are often bought together. Using this data, the store manager can keep the products with high relationships together for display in the store so that the customers might get what they wanted in a single location without searching for them in different locations.

2.Case Study

A case study is done using a general stores dataset.

2.1 Data Set

This is a transactional data set which contains all the transactions that occurred in an wholesale all-occasion gift selling store in the UK. This is a multivariate, Sequential, Time-Series data set with 541909 records and 9 attributes respectively. Data set consists of the following attributes

- Invoice No
- Stock Code
- Description
- Quantity
- Invoice Date
- Unit Price
- Customer
- Country

2.2 Pre-Processing

The dataset needs to be cleaned otherwise it may lead to inconsistent outcome. The dataset which we use consists of around 0.2% of null values in the description column and 24.9% of null values in the Customer ID column. The record with these null values is removed. *Fig. 1* and *Fig. 2* represent the before and after of removing the records with null values.

Certain records in the dataset are duplicates. These records are removed from the dataset. The dataset consists of certain records which represent the orders which are cancelled.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
null values	0.0	0.0	1454.000000	0.0	0.0	0.0	135080.000000	0.0
% of null values	0.0	0.0	0.268311	0.0	0.0	0.0	24.926694	0.0

Fig. 1 Percentage of records with Null value

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
null values	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
% of null values	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 2 Percentage of Null values after removing records

These records are identified by their invoice number which have the letter 'C' followed by the actual order initially placed by the customer. These records are removed along with their counterparts. In certain cases, the records for cancelled orders might not have the counterpart, they are also identified and removed.

2.3 RFM Segmentation

The recency, frequency and the monetary values are extracted from the dataset and stored in a separate data frame (*Fig. 3*). The quartile values of the above attributes are also found with respect to the values 0.25, 0.50, 0.75 (*Fig. 4*).

	Recency	Frequency	Monetary
CustomerID			
12346	325	1	77183.60
12347	39	7	163.16
12348	75	4	331.36
12349	18	1	15.00
12350	310	1	25.20

Fig.3 Recency, Frequency and Monetary value

	Recency	Frequency	Monetary
0.25	21.0	1.0	17.70
0.50	60.0	2.0	47.00
0.75	160.0	5.0	129.73

Fig. 4 R, F, M Quartiles

The RFM score is calculated by using the quartile values. The R-Quartile, F-Quartile and M-Quartile values are initially calculated by finding on which quartile the given value is present. These values are combined together to form the RFM score as shown in *Fig. 5*

	Recency	Frequency	Monetary	R_Quartile	F_Quartile	M_Quartile	RFMScore
CustomerID							
12346	325	1	77183.60	1	1	4	114
12347	39	7	163.16	3	4	4	344
12348	75	4	331.36	2	3	4	234
12349	18	1	15.00	4	1	1	411
12350	310		25.20	1	1	2	112

Fig. 5 RFM score.

The loyalty of the customer is decided based on the RFM score. For this dataset, the threshold for the loyal customers is set to 222. Customers with the RFM score above 222 are said to be the loyal customers and the rest of the customers are not so loyal to the store (Fig. 6). Now the distribution of the loyal and the non-loyal customers is displayed in scatter plot representation (Fig.7).

Ľ→		Recency	Frequency	Monetary	R_Quartile	F_Quartile	M_Quartile	RFMScore	Loyality
	CustomerID								
	12346	325		77183.60			4	114	NotLoyal
	12347	39	7	163.16	3	4	4	344	Loyal
	12348	75	4	331.36	2	3	4	234	Loyal
	12349	18	1	15.00	4	1	1	411	Loyal
	12350	310	1	25.20		1	2	112	NotLoyal

Fig.6 Customer Loyalty data

Now to find the loyalty for future customers, we use a classification model. The Recency, Frequency and Monetary values of the customers are used as feature and the Loyalty as the label for the classification model (*Fig. 8*). The Decision Tree classification algorithm is used to train the model and achieve 99 % accuracy.

Concise the loyal customer into a separate dataframe, which would be used for further processing.



Fig. 7 Scatter Plot for Customer Loyalty

M. Devaganesh et al. / International Research Journal of Multidisciplinary Technovation 2020; 2(3): 14-21



Fig. 8 Decision tree graph for customer loyalty classification

2.3 Product Segmentation

The product description is now used for grouping the products (*Fig. 9*). The important keywords are used for the clustering process. K-Means clustering algorithm is used to group the products. Five clusters of products are formed from the dataset.



Fig. 9 Word Occurence

2.4 Customer Classification

The customers are classified based on the products they purchase. To do this, we need to find the respective cluster the customer prefers the most. This data is now used to develop a

prediction model to predict the product a customer might buy in the future. Random forest classification algorithm is used to train the prediction model.

2.5 Product Association

The products which are bought together in the most transactional cases are found using apriori association algorithm. The support, confidence and lift values are used to identify the rules which are more likely to occur in real life (*Fig. 10*).

[]	rules	[(rules['lift'] >= 6) & (rules['confidence'] >= 0.8)]								
Đ		antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
	2	(ALARM CLOCK BAKELIKE RED)	(ALARM CLOCK BAKELIKE GREEN)	0.104816	0.107649	0.087819	0.837838	7.783073	0.076535	5.502833
	3	(ALARM CLOCK BAKELIKE GREEN)	(ALARM CLOCK BAKELIKE RED)	0.107649	0.104816	0.087819	0.815789	7.783073	0.076535	4.859571
	26	(SET/6 RED SPOTTY PAPER PLATES)	(SET/6 RED SPOTTY PAPER CUPS)	0.135977	0.147309	0.130312	0.958333	6.505609	0.110281	20.464589
	27	(SET/6 RED SPOTTY PAPER CUPS)	(SET/6 RED SPOTTY PAPER PLATES)	0.147309	0.135977	0.130312	0.884615	6.505609	0.110281	7.488196
	28	(ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI	(ALARM CLOCK BAKELIKE GREEN)	0.082153	0.107649	0.070822	0.862069	8.008167	0.061978	6.469547
	29	(ALARM CLOCK BAKELIKE PINK, ALARM CLOCK BAKELI	(ALARM CLOCK BAKELIKE RED)	0.082153	0.104816	0.070822	0.862069	8.224604	0.062211	6.490085
	30	(ALARM CLOCK BAKELIKE GREEN, ALARM CLOCK BAKEL	(ALARM CLOCK BAKELIKE PINK)	0.087819	0.113314	0.070822	0.806452	7.116935	0.060870	4.581209
	41	(SET/6 RED SPOTTY PAPER PLATES, SET/20 RED RET	(SET/6 RED SPOTTY PAPER CUPS)	0.107649	0.147309	0.104816	0.973684	6.609818	0.088958	32.402266
	42	(SET/6 RED SPOTTY PAPER CUPS, SET/20 RED RETRO	(SET/6 RED SPOTTY PAPER PLATES)	0.107649	0.135977	0.104816	0.973684	7.160636	0.090178	32.832861

Fig 10. Association rules generated by Apriori

3. Conclusion

This project detailed various customer classification and market analysis models using the data mining approach. To deal with the recent upgrade in the market our model uses a much more refined analysis on customer purchase history. This model is capable of providing proper insights about the customers to make a useful decision for the successful management of General Stores.

References

- [1] W. Reinartz, V. Kumar, The Mismanagement of Customer Loyalty, *Harvard Business Review*, 80 (2002) 86-94.
- [2] F.F. Reichheld, Loyalty-Based Management, Harvard Business Review, 71 (1993) 64-73.
- [3] A.J. Dubinsky, R.W. Hansen, Improving Marketing Productivity: The 80/20 Principle Revisited, *California Management Review*, 25 (1982) 96-105.
- [4] M. Khajvand, K. Zolfaghar, S. Ashoori, S. Alizadeh, Estimating Customer Lifetime Value Based on RFM Analysis of Customer Purchase Behavior: Case Study, *Procedia Computer Science*, 3 (2011) 57-63.
- [5] Xiang-Bin Yan, Yi-Jun Li, (2006) Customer Segmentation based on Neural Network with Clustering Technique, *In Proceedings of the 5th WSEAS International Conference, on Artificial Intelligence*, Knowledge Engineering and Data Bases, Madrid, Spain, 265-268.
- [6] Ling Luo, Bin Li, Irena Koprinska, Shlomo Berkovsky, Fang Chen, (2017) Tracking the Evolution of Customer Purchase Behavior Segmentation Via A Fragmentation-Coagulation Process, Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI), 2414-2420.

M. Devaganesh et al. / International Research Journal of Multidisciplinary Technovation 2020; 2(3): 14-21

- [7] D. Chen, S. L. Sain, K. Guo, Data Mining for The Online Retail Industry: A Case Study of RFM Model-Based Customer Segmentation Using Data Mining, *Journal of Database Marketing & Customer Strategy Management*, 19 (2012) 197-208.
- [8] Y. Zhang, J. Jiao, Y. Ma, Market Segmentation for Product Family Positioning Based on Fuzzy Clustering, *Journal of Engineering Design*, 18 (2007) 227-241.
- [9] Narendra Kumar Jha, Manoj Kumar, Anurag Kumar, Vijay Kumar Gupta, Customer Classification in Retail Marketing by Data Mining, *International Journal of Scientific & Engineering Research*, 5 (2014) 325-330.
- [10] Y. Zhang, J. Jiao, Y. Ma, Market Segmentation for Product Family Positioning Based on Fuzzy Clustering, *Journal of Engineering Design*, 18 (2007) 227-241.
- [11] A.A. Raorane, R.V. Kulkarni, B.D. Jitkar, Association Rule-Extracting Knowledge Using Market Basket Analysis, *Research Journal of Recent Sciences*, 1 (2012) 19-27.

Funding

This study was not funded by any grant

Conflict of interest

None of the authors have any conflicts of interest to declare.

About The License

The text of this article is licensed under a Creative Commons Attribution 4.0 International License

Cite this Article

M. Devaganesh, B. Dilipan, R. Abirami, Customer Relationship and Sales Management using Data Mining Techniques, International Research Journal of Multidisciplinary Technovation, Vol 2, Iss 3 (2020) 14-21.

DOI: https://doi.org/10.34256/irjmt2033