SHM PUBLISHER
SURYA HIJAU MANFAAT

Journal of Information System Exploration and Research

https://shmpublisher.com/index.php/joiser

p-ISSN 2964-1160 | e-ISSN 2963-6361

Implementation of Loyalty Program Theory Based on Recency Frequency Monetary Score in Information Systems to Increase Customer Loyalty

Ricky Mangihut Rajagukguk[1*]

[1]*Department of Computer Science, Universitas Negeri Semarang, Indonesia*

DOI: https://doi.org/10.52465/joiser.v3i1.538

Received 28 January 2025; Accepted 30 January 2025; Available online 31 January 2025

Article Info

Keywords:

RFM score;
K-Means clustering;
Loyalty program;
Customer segmentation;
Online retail;

Abstract

This study aims to help online retail stores find the right strategy for treating customers through customer segmentation based on Recency, Frequency, and Monetary (RFM) Score. With a quantitative approach, this study uses the K-Means Clustering algorithm to group customers based on their RFM values and applies it within the Loyalty Program Theory framework. The results show that the Best Customers segment has the highest percentage at 26.3%, which emphasizes the importance of retaining high-value customers through exclusive loyalty programs such as VIP access and premium offers. In contrast, the Lost Customers segment at 24.8% requires attention through retargeting and discount programs to attract them back. This study proves that data-based customer segmentation and the implementation of relevant strategies can strengthen long-term relationships with customers, increase loyalty, and ultimately help the development of online retail businesses.

1. Introduction

The world that has been increasingly developing especially in the field of digitalization until now has made some significant changes to society, especially in consumption patterns, one example is in the field of online retail sales which is the main subject of this study. Online retail is a way of selling products online that involves selling goods directly to end consumers for personal use [1]. E-commerce has become one of the dominant business models, where consumers can easily make transactions online, this is reinforced by Anggaranie in 2017 [2]. E-commerce has changed the classical business paradigm by creating interaction models between producers and consumers in the virtual world.

Recency, Frequency, and Monetary, commonly abbreviated as RFM is a popular method for analyzing customer behavior for more than 50 years [3]. The RFM model is a behavior-based model used to analyze customer behavior and predict customer loyalty based on database behavior [4]. As

[*] ***Corresponding Author:***

Ricky Mangihut Rajagukguk,
Department of Computer Science,
Universitas Negeri Semarang,
Semarang, Indonesia.
Email: kyy@students.unnes.ac.id

the name suggests, this model is divided into 3 factors, namely how often they make purchases (Recency), how often they make transactions (Frequency), and how much money (Monetary) they spend.

The algorithm used to divide customers according to their respective RFM behavior is the K-Means Algorithm. In the K-means method, data is divided into several groups, where one group has identical characteristics, but has different characteristics from other groups [5]. With K-Means, companies can identify clusters of customers who share similar characteristics in terms of purchase frequency, transaction value, and last transaction time. However, determining the right number of clusters (K) is a crucial step in generating accurate segmentation. This process usually involves evaluation using methods such as the Elbow Method that helps in identifying the optimal K value.

After segmenting customers based on RFM analysis, the next step is to implement an effective strategy to increase customer loyalty. One approach that can be used is the Loyalty Program Theory, which focuses on developing long-term relationships with customers by providing incentives such as points, discounts, or exclusive services. This kind of loyalty program will greatly help businesses retain customers, increase satisfaction, and prevent customers from being tempted by various offers offered by other competitors [6]. By leveraging RFM segmentation results, companies can design more targeted offers, such as providing special incentives to loyal customers or implementing retargeting programs for customers who are about to lose interest.

2. Method

This study uses a quantitative approach where numerical data such as Recency, Frequency, Monetary, Frequency, and Monetary are analyzed to obtain RFM scores as a benchmark for grouping customers. According to Musianto in 2002 [7] quantitative approach uses aspects of measurement, calculation, formulas, and data certainty in the form of numbers. The results of this study will be in the form of a statistically measurable customer classification so that it can support strategic decision-making based on data.

The dataset used in this study is an online retail sales dataset taken from the site www.kaggle.com. Kaggle is a site that provides various datasets to be processed using machine learning, the license on this dataset is general so that users are free to process it [8]. The next step is to calculate the RFM score of each customer. This RFM score is used to understand customer transaction behavior and provide a strong basis for segmentation. Each customer will be given an R, F, and M score to form a profile that will be used in the clustering process.

The data that has been collected will then be analyzed using the Rapid Miner software tool, with the application of the K-means algorithm to facilitate the research process. Rapidminer is a data science software developed by a company of the same name that provides an integrated environment for machine learning, deep learning, text mining, and predictive analytics [9].

To find the optimal K value, the elbow method is needed or commonly called the elbow method. The elbow method is widely used because this method can determine the number of clusters well, and the results of determining a good K value can be used to maximize cluster results [10]. The way the elbow method works is after getting the average distance results from K = 1 to K = N, then the results will be applied to the statistical model of the line diagram, which then looks at which part or point forms the elbow, that point will be the optimal K value.

Loyalty Program Theory is applied in this study to provide relevant strategies based on customer segmentation results using RFM scores. This approach aims to improve long-term relationships with customers by providing appropriate incentives. According to Holbrook and Chaudhuri in 2001 [11]. Loyalty programs can increase behavioral loyalty and attitudinal loyalty through strategies designed based on customer preferences and behavior.

3. Results and Discussion

The results and discussion of this study will discuss the findings generated based on the methodology that has been described previously. Through a series of research steps carried out, the author managed to collect relevant data, analyse it in depth, and provide an in-depth interpretation of the results obtained.

3.1. Dataset

The following is the dataset used in this research which contains online retail store sales taken from Kaggle.

Table 1. Sales dataset

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01/12/2010 08:26	2,55	17850	United Kingdom
536365	71053	WHITE METAL LANTERN	6	01/12/2010 08:26	3,39	17850	United Kingdom
536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01/12/2010 08:26	2,75	17850	United Kingdom
536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01/12/2010 08:26	3,39	17850	United Kingdom
This dataset consists of 4336 rows.							

3.2. Pre-processing

This pre-processing section involves a series of steps aimed at cleaning, preparing, and converting data into a format that is easier to understand and ready to be processed. These cleaning steps are applied to improve data processing results and minimize error rates. After data pre-processing, several columns are removed, namely columns other than numeric types, and because more than 90% of customers are United Kingdom citizens, this study only focuses on customers who are United Kingdom citizens.

Table 2. List of the countries

Country	Count	Ratio
United Kingdom	480471	91.48%
Germany	8998	1.71%
France	8376	1.59
And the other		

3.3. RFM modeling

In this stage, the dataset that has gone through the pre-processing process is transformed to calculate the RFM model for each customer, can be seen in Table 3.

Table 3. RFM model

CustomerID	*Recency*	*Frequency*	*Monetary*
12346.0	326	1	263.12
12347.0	3	7	4299.8
12348.0	76	4	1797.24
….	….	….	….
18282.0	8	2	178.05
18283.0	4	16	2094.88
18287.0	43	3	1837.28

The RFM value for each customer will later be converted into an RFM Score and the RFM Score will provide an overview of the loyalty and potential of each customer, which can be used for a more targeted marketing strategy. Table 4 is the RFM Score.

Table 4. RFM score

CustomerID	*Recency*	*Frequency*	*Monetary*	RFM Score
12346.0	2	1	2	21
12347.0	3	4	5	34
12348.0	3	4	5	34
….	….	….	….	….
18282.0	5	3	1	53
18283.0	5	5	5	55
18287.0	3	3	4	33

This modeling is an important initial step before proceeding to the segmentation stage using the K-means algorithm, which can group customers into more specific categories based on identified behavioral patterns.

3.4. Optimal K-value

The next step is to find the optimal K-value. In finding the optimal K-value using Rapid Miner software as a tool to help find everything needed.

3.4.1. Clustering

Aims to divide the dataset into several clusters that can be arranged according to user wishes.

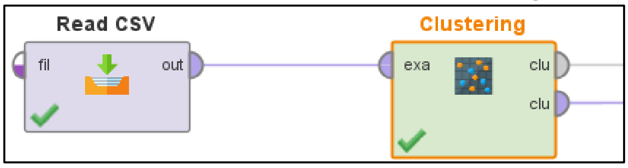

Figure 1. Clustering

Following the previous explanation, the number of clusters (K) can be adjusted according to the user's wishes, the method is to open the parameters of the "clustering" operators, then set the number of "K" by entering a number according to the desired number of clusters to be divided.

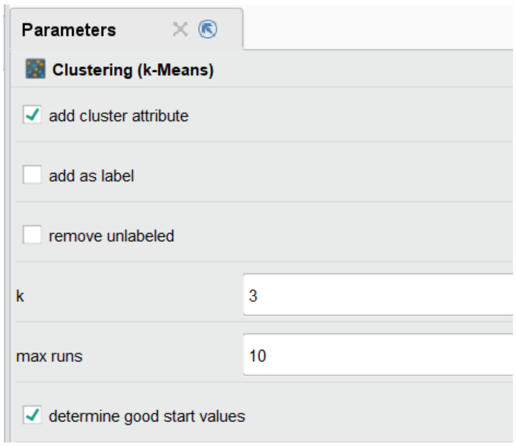

Figure 2. Clustering parameter

The following is an example of the results of successfully dividing the dataset into three clusters. From a total of 4,319 record, it was found that the data was divided into the third cluster (Cluster_2).

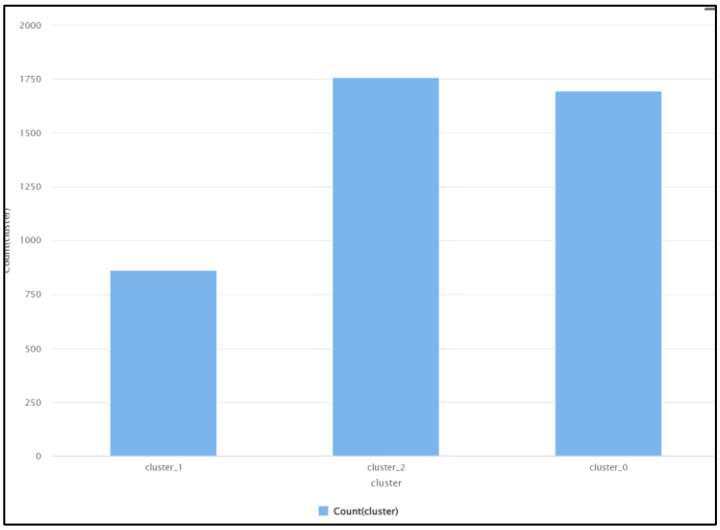

Figure 3. Clustering result K = 3

3.4.2. Average Distance Performance

Because the elbow method requires the results of Average Distance Performances, the operator used is "Cluster distance performance". ADP is one of the most common methods used to determine the optimal k value in k-means clustering. Drag to the process section, then connect the operators with the dataset and other operators as shown in Figure 4.

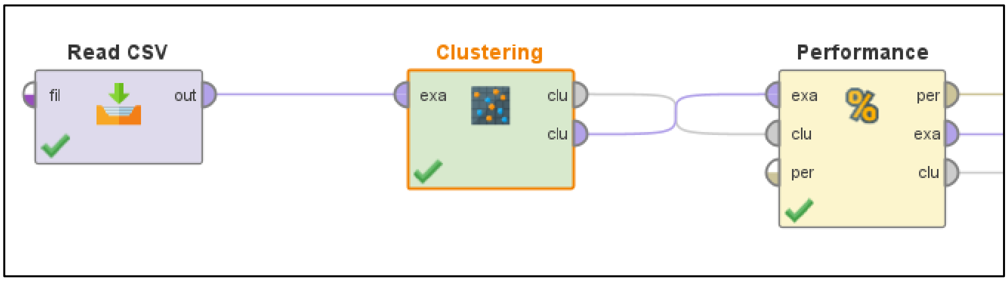

Figure 4. Average distance performance

Just like with the previous operators, set the parameters of the "Cluster distance performance" operator according to the image below, to produce the correct value.

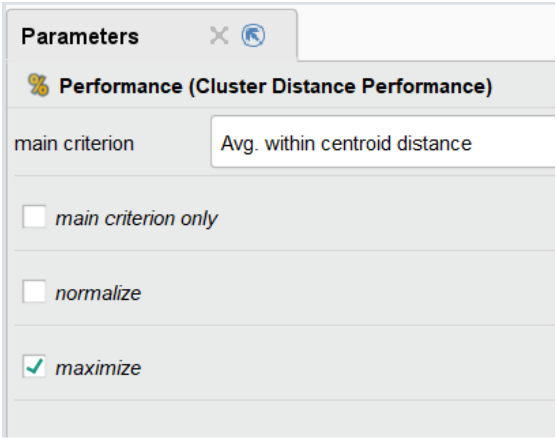

Figure 5. Parameter Average Distance performance

The purpose of checking the maximize section is so that the value of the average distance performances is positive, the result of the Average Distance Performances of the three clusters is 23,248. The meaning of the value of 23,248 is that on average, the data points in the cluster are about 23,248 units (or units used in the data scale) from the nearest cluster center. The smaller the average distance value, the denser or more homogeneous the cluster is, which indicates that the data points in the cluster are well adjacent to each other. To find the optimal K value, several tests are needed several times with different K values, in this case, I will do testing with K = 2 to K = 7. And here are the results of the testing that has been done.

Table 5. Result of Average Distance Performance Test

K	Average Distance Performance
2	56.606
3	23.248
4	12.549
5	1.449
6	1.184
7	0.945

3.4.3. Elbow method

The final step in finding the optimal K-value is the elbow method. This method is a method for determining the optimal number of clustering by looking at the K-value points that resemble a right angle. The elbow method is a simple and useful approach to determining the optimal number of clusters in clustering. However, they also note that the interpretation of the elbow graph is not always clear, and can sometimes be subjective. The results are shown in Figure 6.

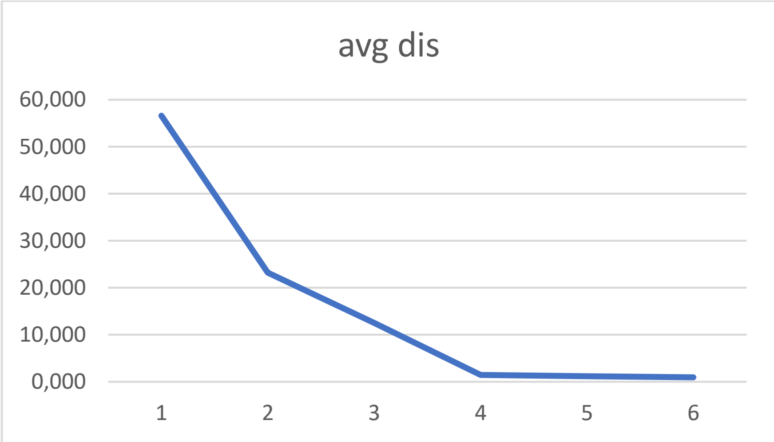

Figure 6. The result of the elbow method

Based on the line diagram graph above, the K value point that is more like a right angle is at point 4, which means that K = 5 is the optimal number of clusters. By finding the optimal number of clusters, it can help improve the efficiency of data grouping. By determining the optimal number of clusters, decisions based on cluster analysis become more accurate and efficient. This is because the optimal number of clusters reflects the most relevant and meaningful data separation [12].

3.5. Cluster Division Using K-Means

Since it is known that the optimal K value is five, the cluster division is according to Table 6.

Table 6. Cluster division

RFM score	Segment	Characteristic
11-22	Lost Customers	R↓F↓M↓
23-33	At-Risk Customers	R—F↓M↓
34-41	Promising Customers	R—F—M—
42-51	Loyal Customers	R↑F—M—
52-55	Best Customers	R↑F↑M↑

This segmentation table illustrates the division of customers based on RFM (Recency, Frequency, Monetary) scores with the following characteristics: Lost Customers **(R↓ F↓ M↓)** indicates inactive, infrequently transacting, and low-spending customers. At-risk customers **(R— F↓ M↓)** have better Recency but still have low frequency and expenses. Promising Customers **(R— F— M—)** indicates customers who are active but do not yet have high frequency or spending. **Loyal Customers (R↑ F— M—)** are customers who frequently make transactions with moderate spending values, and **Best Customers (R↑ F↑ M↑)** are very loyal customers with high transactions and large expenses.

After naming each cluster, here are the results of clustering mapping for each online retail store customer which can be seen in Table 7.

Table 7. Final result

CustomerID	Recency_Score	Frequency_Score	Monetary_Score	RFM Score	Segment
16161	5	5	5	55	Best Customers
15555	5	5	5	55	Best Customers
16954	5	5	5	55	Best Customers
….	….	….	….	….	….
14889	1	1	1	11	Lost Customers
13747	1	1	1	11	Lost Customers
14589	1	1	1	11	Lost Customers

The grouping results above show that the "Best Customers" cluster consists of 1,139 customers, the "Loyal Customers" cluster consists of 889 customers, the "Promising Customers" cluster consists of 581 customers, the "At-Risk Customers" cluster consists of 656 customers, the "Lost Customers" cluster consists of 1,071 customers. The image below shows that the "Best Customer" cluster has the highest percentage with 26.3% while the "Promising Customer" cluster has the lowest percentage with 13.4%.

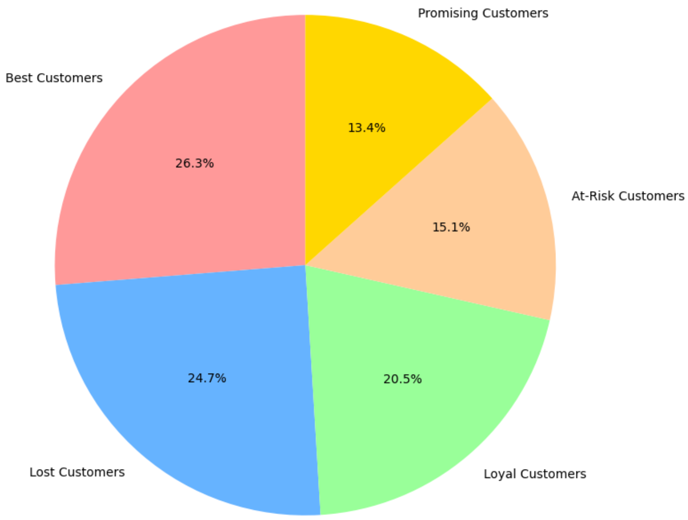

Figure 7. Customer distribution by segment

3.6. Loyalty Program Theory

The appropriate strategy based on the results of the customer segmentation division is that in the Lost Customers segment, the company must attract back customers who have not made transactions for a long time through attractive offers or retargeting programs, which is in line with the theory [13] about the importance of restoring relationships with lost customers. For At-Risk Customers, customers who have the potential to stop trading need to be given special attention with exclusive offers or promotional campaigns to prevent them from leaving the company, according to the concept discussed by [14] regarding the importance of maintaining at-risk customers. In the Promising Customers segment, the company focuses on strengthening relationships with customers who show the potential to become loyal customers, through product recommendations and special offers, which supports the concept proposed by [14] and [15]. Loyal Customers need loyalty incentives, such as vouchers or reward points, to strengthen their long-term relationship with the company, which is in line with the theory [13] and [14] regarding the awarding of rewards to maintain loyalty. As for Best Customers, who are the most valuable customers, the company gives more attention to exclusive services such as VIP access or premium offers, by the principles of loyalty theory that prioritize rewards for the best customers to maintain their loyalty, as explained by [15].

4. Conclusion

This research shows that the application of Loyalty Program Theory through RFM Score-based customer segmentation (Recency, Frequency, Monetary) can improve the efficiency of marketing strategies by providing a more targeted approach to each customer segment. Using the K-Means Clustering algorithm, customers are grouped into five main segments: Lost Customers, Risky Customers, Promising Customers, Loyal Customers, and Best Customers, each with different characteristics and marketing strategies. The findings revealed that the Best Customers segment had the highest percentage (26.3%), indicating the importance of retaining customers with high transaction values through exclusive loyalty programs such as VIP access and premium offers. On the other hand, the Lost Customers segment (24.8%) requires special attention through retargeting and discount programs to attract them back. This study confirms that data-driven customer segmentation and the implementation of relevant strategies by loyalty theory can enhance long-term relationships with customers, strengthen loyalty, and ultimately increase company profitability.

References

[1] B. S. Ashari, S. C. Otniel, and Rianto, 〝Perbandingan Kinerja K-Means Dengan DSCAN Untuk Metode Clustering Data Penjualan Online Retail,〞 *J. Siliwangi*, vol. 5, no. 2, pp. 72−77, 2019.

[2] G. Anggaranie, 〝Perkembangan E-Commerce Beserta Klasifikasinya,〞 *J. SupplyChain*, pp. 1−4, 2017.

[3] A. Sciences, 〝CUSTOMER SEGMENTATION BY USING RFM MODEL AND CLUSTERING METHODS : A CASE,〞 pp. 1−19, 2018.

[4] Y. H. Chrisnanto and A. Kanianingsih, 〝Pengelompokan Ekuitas Pelanggan Berbasis Recency Frequency Monetary (Rfm) Menggunakan K-Means Clustering,〞 *Semin. Nas. Teknol. Inf. dan Komun. 2019 (SENTIKA 2019)*, vol. 2019, no. Sentika, 2019.

[5] P. Apriyani, A. R. Dikananda, and I. Ali, 〝Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi,〞 *Hello World J. Ilmu Komput.*, vol. 2, no. 1, pp. 20−33, 2023, doi: 10.56211/helloworld.v2i1.230.

[6] H. K. Sari, 〝Efektivitas Loyalty Program dalam Customer Relationship Management terhadap Kepuasan dan Loyalitas Pelanggan,〞 *Ilmu Komun.*, p. 30, 2009.

[7] L. S. Musianto, 〝Perbedaan Pendekatan Kuantitatif Dengan Pendekatan Kualitatif Dalam Metode Penelitian,〞 *J. Manaj. dan Wirausaha*, vol. 4, no. 2, pp. 123−136, 2002, doi: 10.9744/jmk.4.2.pp.123-136.

[8] A. . Rahmat, M. . Ladjamuddin, and T. . Awaludin, 〝Perbandingan Algoritma Decision Tree, Random Forest Dan Naive Bayes Pada Prediksi Penilaian Kepuasan Penumpang Maskapai Pesawat Menggunakan Dataset Kaggle,〞 *J. Rekayasa Inf.*, vol. 12, no. 2, pp. 150−159, 2023.

[9] R. Nofitri and N. Irawati, 〝Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer,〞 *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 199−204, 2019, doi: 10.33330/jurteksi.v5i2.365.

[10] D. A. I. C. Dewi and D. A. K. Pramita, 〝Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,〞 *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102−109, 2019, doi: 10.31940/matrix.v9i3.1662.

[11] A. Chaudhuri and M. B. Holbrook, 〝The Chain of Effects from Brand Trust and Brand Affect to Brand Performance: The Role of Brand Loyalty,〞 *J. Mark.*, vol. 65, no. 2, pp. 81−93, Apr. 2001, doi: 10.1509/jmkg.65.2.81.18255.

[12] D. J. KETCHEN and C. L. SHOOK, 〝THE APPLICATION OF CLUSTER ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH: AN ANALYSIS AND CRITIQUE,〞 *Strateg. Manag. J.*, vol. 17, no. 6, pp. 441−458, Jun. 1996, doi: https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.

[13] A. S. Dick and K. Basu, 〝Customer loyalty: Toward an integrated conceptual framework,〞 *J. Acad. Mark. Sci.*, vol. 22, no. 2, pp. 99−113, 1994, doi: 10.1177/0092070394222001.

[14] B. Sharp and A. Sharp, 〝Loyalty programs and their impact on repeat-purchase loyalty patterns,〞 *Int. J. Res. Mark.*, vol. 14, no. 5, pp. 473−486, 1997, doi: https://doi.org/10.1016/S0167-8116(97)00022-0.

[15] V. Kumar and D. Shah, 〝Building and sustaining profitable customer loyalty for the 21st century,〞 *J. Retail.*, vol. 80, no. 4, pp. 317−329, 2004, doi: https://doi.org/10.1016/j.jretai.2004.10.007.