

CUSTOMER SEGMENTATION USING DATA MINING TECHNIQUES: AN ANALYSIS OF ONLINE RETAIL INDUSTRY

Kulwant Singh, Research Scholar, Department Of Computer Science, Shri Khushal Das University, Hanumangarh (Rajasthan)-India

Dr. Garima Bansal, Research Supervisor, Department Of Computer Science, Shri Khushal Das University, Hanumangarh (Rajasthan)-India

Abstract

The goal of this study is to determine consumer purchasing patterns, create customer segments, and determine client addresses for Bukku.id. This study makes use of Bukku.co.id's consumer purchase information from September 1 through September 17, 2022.2023. To determine consumer segmentation, the RFM technique and clustering are employed. The results of the Pareto analysis indicate which efforts should be prioritized by publishers and authors to reap the most rewards. In order to decide promotion and offline marketing plan, customer address or location has been mapped based on priority authors. The study's findings, which are based on RFM and clustering analysis, identify three customer clusters. The grouping method is a critical stage in the data mining process. A multivariate methodology is great for market gauging and arranging research segmentation applications. This study paper gives an intensive investigation of the k-implies bunching strategy utilizing the SPSS Instrument to make an online, ongoing framework for a particular grocery store to estimate deals during different yearly occasional cycles.

Keywords: Customer Segmentation,Data Mining,Online,Retail

1. INTRODUCTION

Highlight A statistical method very similar to classification is clustering. It groups reasonably uniform observations and significant groupings of raw data. While the objects in one cluster share some traits and features with those in others, they do not. Finding commonalities between data based on traits discovered in raw data allows for grouping. The major goal was to determine the ideal number of clusters. Hierarchical and non-hierarchical clustering methods are the two main categories. The process of clustering is an ongoing, iterative method of extracting knowledge from vast amounts of unstructured, raw data. To achieve the best results, the right clustering technique and parameters must be chosen for a specific classification task. Exploratory data mining techniques like clustering are employed in a variety of application-oriented fields like machine learning, classification, and pattern recognition.

Data mining has recently accelerated significantly for knowledge-based services like distributed and grid computing.

The most crucial characteristic of the clustering method is that a tuple from one cluster is more likely to be similar to other tuples from the same cluster than tuples from other clusters. The interchange of products and services depends heavily on the retail industry. The retail sector serves as a hub for the direct interaction of various producers and wholesalers with the final consumer. In addition to distributing items to customers, they offer other crucial services. Customers are increasingly looking for retailers who can deliver great services, and they are drawn to attractive establishments with a wide selection of goods and point of sale (POS). Every retailer is concerned with developing strategies to draw in more customers in order to boost sales. Without a question, retail centers with efficient service systems will draw in more patrons. This should include a strategic plan for making it simple for customers to pay for things, find them on the shelf, and collect them. For this reason, it is simple to state that services provided by the retail sector are frequently customized to meet the needs of the clientele. Customers behave differently from one another as a result of differences in their shopping habits, shifting needs, shifting desires, and inconsistent brand preferences. Keeping track of

data on client behavior is crucial for making decisions about how to provide excellent customer service.

2. REVIEW OF LITREATURE

MSE Kasem 2023 Effective communication between marketing and senior management is essential for success in a company climate that is becoming more and more customer-centric. Utilizing modern data mining tools to find potential customers is crucial for direct marketing operations in light of the development of globalization and greater competition. In order to construct a customer profiling system and enhance sales performance, this study suggests a data mining preprocessing method that includes customer equity estimation and customer action prediction. To assess customer capital, the RFM-analysis methodology is applied, and a boosting tree is used for forecasting.

L Rajput, SN Singh - 2023 To help firms maximize their investment, a variety of data mining techniques, including classification, clustering, regression, etc., are used to analyze client purchase patterns. In order to determine whether to prioritize its website or mobile application, an e-commerce company's real-time data set is clustered in this study. Due to the dispersed nature of the data, the K-means clustering technique is employed to partition and cluster the users in order to identify hidden patterns in the data set.

S Peker, 2023 The amount of customer transactional data that has been made available to businesses has grown in recent years as digital transformation has gained momentum. Customer segmentation has drawn significant attention from various industries thanks to the use of such enormous amounts of transactional data and the application of various data mining techniques. Significant research effort has also been devoted to this topic, and the body of literature has started to grow.

The paper presents an experimental study on customer segmentation for personalized marketing in online retailing. The authors propose an approach that combines RFM analysis and K-means clustering to segment customers based on their recency, frequency, and monetary value. The experiments were conducted using grey images retrieved from MATLAB media or downloaded from the Berkeley image segmentation database. The performance of the proposed approach was compared with two other methods: a histogram-based method and the subtractive clustering method. The evaluation of the segmented image quality was done using the Q-value criterion function. The results showed that the proposed improved K-means method outperformed the other approaches in terms of segmentation quality. The computational time of the proposed method was longer in some cases but provided better results. The discussion also included comparisons with other clustering algorithms, such as FCM and K-medoids, showing that the proposed approach performed well in terms of segmentation quality and computational cost. Overall, the paper presents a promising approach for customer segmentation in online retailing using RFM analysis and K-means clustering, with potential applications in personalized marketing.

The literature review by Liu, Lu, Chen, and Yu (2021) focuses on customer segmentation based on purchasing behavior analysis using data mining techniques in e-commerce. The authors highlight the changing economics of customer relationships due to globalization, internet and telecommunication networks, and information technology. They emphasize the importance of analyzing customer lifecycle, understanding the customer's psychological mindset, and utilizing data mining, customer relationship management (CRM), and other tools to enhance relationship marketing. The review discusses the shift from a product-oriented view to a customer-oriented view, the role of data mining in uncovering hidden patterns and predicting individual behavior, and the impact of the internet on marketing focus. The paper also delves into the concepts of data mining, CRM, customer identification, attraction, retention, and development, as well as various data mining techniques such as association, classification, clustering, forecasting, regression, visualization, and web mining. The authors emphasize the significance of web mining in internet marketing and customer relationship management strategies.

The study by Park and Kim (2022) focuses on enhancing customer segmentation in online retailing using RFM analysis and clustering. The research methodology consists of three major steps: pre-analysis efforts for data cleaning and transformation, data analysis using RFM analysis, two-step cluster analysis, and K-means clustering, and presentation of the results. The study utilizes a secondary dataset obtained from customer loyalty card accounts in a sports retailing company. The dataset includes variables such as recency, frequency, and monetary value (RFM) of customer purchases. The results of the analysis reveal different customer clusters based on their RFM scores. The study compares the proposed customer segmentation with the company's current segmentation based on card types. The findings suggest that the proposed models could lead to changes in the current card type assignments, with a significant percentage of customers potentially being reclassified. The study highlights the importance of understanding customer behavior and using data-driven approaches for effective customer segmentation in online retailing.

3. RESEARCH METHDOLOGY

3.1 Population and Sampling Size

The demographic of online buyers used in this study is sampled so that they are employed and have previously made an online purchase.

For the research, ten different occupational categories are defined. These Businessperson, academic, doctor, dentist, teacher, engineer, attorney, accountant/financial adviser, craftsman, research assistant, and economist are characterized as occupational groups. A questionnaire is used to collect the raw data. The sampling strategy is random sampling. Out of the 200 questionnaires that were issued, 100 have been returned. The questionnaire includes 50 demographic questions (gender, marital status, age, etc.), 20 questions about internet use and behavior (using the internet, the most important reason for doing so, how often you shop online, how many hours you spend online each day, etc.), and 30 questions about your online shopping habits Concerning Smith and Swinyard's internet-related lifestyle scale, particular focus is given to the issue of construct equivalence and translation equivalent, and such issues are reduced. A five-point Likert scale is used to evaluate the respondents' lifestyle traits; the responses 1, 2, 3, 4 and 5 stand for "I strongly disagree," "I disagree," "I have no idea," "I agree," and "I strongly agree," respectively. Table 1 below lists the sociodemographic and internet usage characteristics of the respondents.

4. ANALYSIS AND RESULT

Table 1:Online shoppers' sociodemographic and internet usage characteristics (n=200)

Items	Frequency	Percentage
Gender		
Male	120	60
Female	80	40
Age		
20-30	60	30
30-40	50	25
40-50	50	25
50 above	40	20
Education		

Primary School	30	15
Middle School	40	20
High School	50	25
B. Sc.	25	12.5
M.sc	25	12.5
Phd	30	15
Marital Status		
Single	80	40
Married	60	30
Widowed	60	30
Number of Household		
1	50	25
2	20	10
3	30	15
4	40	20
5	60	30

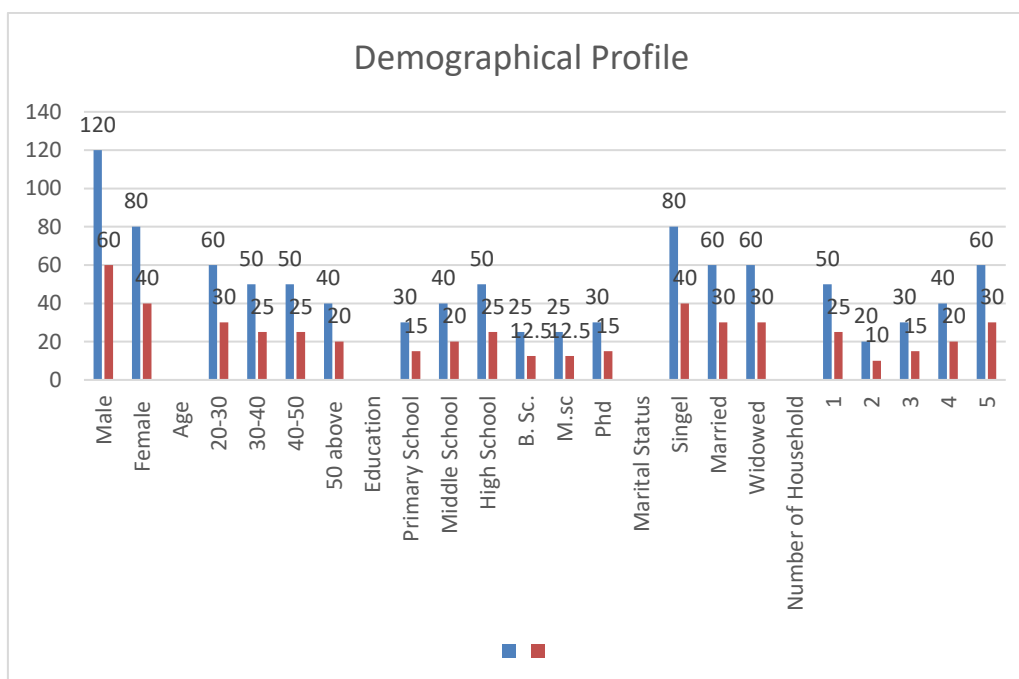


Figure 1: Online shoppers' sociodemographic and internet usage characteristics (n=200)

4.2 Steps of the Model

The segmentation model used in this case, Clementine 8.1, works with data streams. The hubs that are associated with each other with bolts exploit each procedure on the stream. A source hub, a

field/record hub, and a result hub make up an essential stream. In Figure 2, the segmentation model is shown.

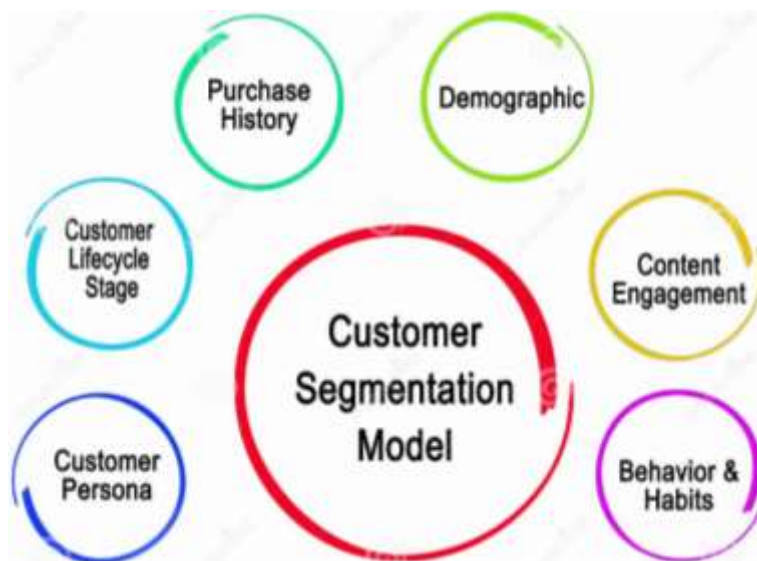


Figure 2: The Segmentation Model

The model begins with a "database" node as shown in Figure 1 and continues with a "type" node, which is used to specify the field kinds.

A "Kohonen" modeling node is added to the "type" node following the specification of the field types. Then each item's "Kohonen" node is executed. The "Kohonen" node is used to identify eight clusters. Table 2 displays the top three clusters and their respective characteristics.

These clusters are then used to create segments based on internet-related lifestyle traits by applying "select" nodes to each cluster's entries. The "select" nodes are linked to the "apriori" modeling nodes.

Table 2:Results of Korhonen Modelling

Most Important Attributes	Cluster 1 (X = 0, Y = 0)	Cluster 2 (X = 0, Y = 2)	Cluster 3 (X = 3, Y = 2)
Age	20-30	20-30	3.-40
	30-40		40-50
	40-50		
Most Important Reason	Shopping	Business	Business E-Mail
Education	Bachelor of Science	Master of Science	High School Bachelor of Science
	Master of Science		
	Doctor of Philosophy		
Gender	Male	Male	Male
	Female	Female	
Hours Online Per Day	2-3 Hours	More than 4 Hours	Less Than 2 Hour 2-3 Hours
Internet Usage	More Than 5 Years	More Than 5 Years	More Than 5 Years
Marital Status	Married	Married	Married
	Single	Single	Single
Monthly Income*	2,000-4,000 TRY 4,001-6,000 TRY	Less Than 2,000 TRY	More Than 6,000 TRY
Number of Household	1	1	More than 4
	2	2	
Occupational Group	Engineer	Research Assistant	Businessperson
	Lawyer		

The statements measuring the respondents' internet-related lifestyle traits are stated as antecedents and other elements (such as demographics and internet use) in each "Apriori" modeling node(Behavior) as a result. Various association rules are extracted when each "Apriori" modeling node is completed.

Rule accuracy, often known as precision in information retrieval, is the most significant indicator of rule quality Support, confidence, and lift are the three dimensions of an association rule that provide information about the rule's accuracy. Table3 provides the support, confidence, and lift parameter values calculated for this investigation.

Table 3: Analysis Accuracy

Association Rules	Support (%)	Confidence (%)	Lift
Rules 1	31.2	100.0	1.897
Rule 2	18.6	100.0	1.897
Rules 3	19.8	92.0	1.789

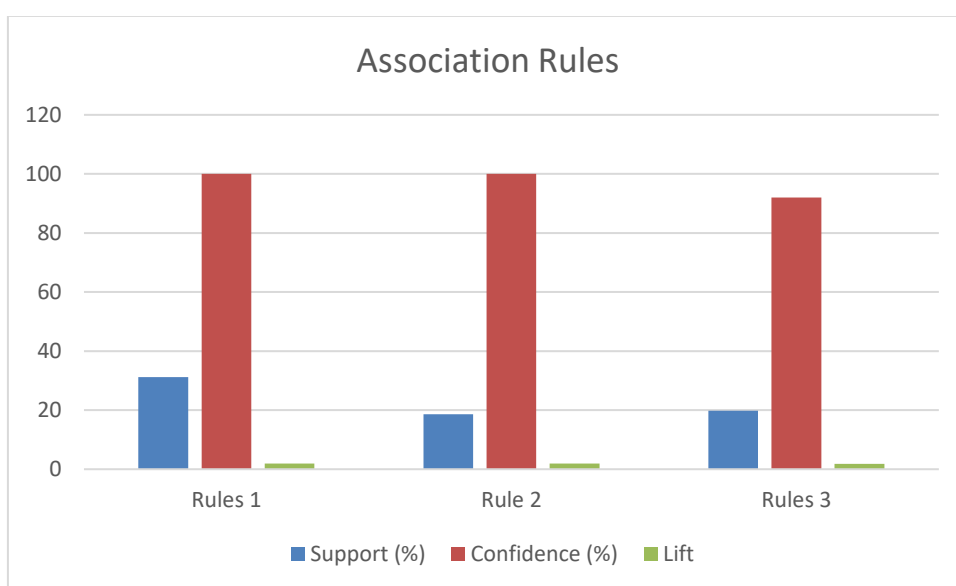


Figure 3: Analysis Accuracy

5. DISCUSSION

The part inside the "1000-3000 Attempt" month to month pay range emphatically leans toward online purchasing, as shown in Table 3 for the extricated affiliation rule in the main group where the ensuing is "month to month pay". "Online shopping darlings" can be utilized to depict this gathering. They appreciate purchasing online since, as displayed in Table 2, it is their essential inspiration for utilizing the web. They accept that online retailers have a more extensive choice than neighborhood shops. They are better taught, have higher PC education, and abhorrence the burden of doing their shopping locally. The subsequent group incorporates one more fragment made up totally of female respondents. This fragment's way of life qualities vary somewhat from those of the first. As indicated by these respondents, the greatest negatives of web buying are transporting costs, costly costs, and security concerns. That's what they note assuming the costs were lower they would buy online more, yet they additionally accept that social collaboration with others is urgent.

The expression "social online customers" may be utilized to depict this gathering. The last affiliation rule in Table 3 connects with male financial specialists' "secondary school" training and absence of trust in the security of online exchanges. They find it trying to utilize and grasp the online requesting process. In contrast with different parts, this one is affluent. These respondents can be alluded to as "dubious online customers" since they have the most negative perspectives on online purchasing.

The examination's discoveries share a few equals with those of studies led by Brengman et al. (2003) and Allred et al. (2006). The discoveries of this concentrate additionally demonstrate that respondents with more significant levels of schooling seem, by all accounts, to be less stressed with security issues, and men have all the earmarks of being more worried about security issues, as found in Tables 2 and 3. Furthermore, female respondents make up the amiable online customer section. Comparable results were found in Hui and Wan's (2006) work. There are continuous stresses over the wellbeing of online purchasing, which are applicable for the "dubious customers" portion distinguished here, as per Lokken et al.

6. CONCLUSION

The reason for this study was to order online purchasers into bunches in view of variables associated with their online ways of life. To help the administration of online buying, the segmentation cycle was finished with the guide of an incorporated data mining methodology that utilizes a self-sorting out map brain organization and affiliation rule mining. In view of web associated way of life highlights, data mining results introduced here show that three special sections connected with month to month pay, orientation, and training still up in the air.

Albeit most of earlier examinations on online customer segmentation have been centered around segment and financial factors, their discoveries have not completely upheld the advertising methodologies of online organizations. Subsequently, an essential descriptor — the web related way of life characteristics of online customers — has gotten little consideration in the exploration.

REFERENCES

1. Kasem, M. S., Hamada, M., & Taj-Eddin, I. (2023). Customer Profiling, Segmentation, and Sales Prediction using AI in Direct Marketing. arXiv preprint arXiv:2302.01786.
2. Rajput, L., & Singh, S. N. (2023, January). Customer Segmentation of E-commerce data using K-means Clustering Algorithm. In 2023 13th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 658-664). IEEE.
3. Peker, S., & Kart, Ö. (2023). Transactional data-based customer segmentation applying CRISP-DM methodology: A systematic review. *Journal of Data, Information and Management*, 1-21.
4. Samli, A. C., Pohlen, T. L., & Bozovic, N. (2002). A review of data mining techniques as they apply to marketing: Generating strategic information to develop market segments. *The Marketing Review*, 3(2), 211-227.
5. Yıldız, E., GüngörŞen, C., & Işık, E. E. (2023). A Hyper-Personalized Product Recommendation System Focused on Customer Segmentation: An Application in the Fashion Retail Industry. *Journal of Theoretical and Applied Electronic Commerce Research*, 18(1), 571-596.
6. Sun, Y., Liu, H., & Gao, Y. (2023). Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model. *Heliyon*, 9(2).
7. Sivaguru, M. (2023). Dynamic customer segmentation: a case study using the modified dynamic fuzzy c-means clustering algorithm. *Granular Computing*, 8(2), 345-360.
8. Griva, A., Zampou, E., Stavrou, V., Papakiriakopoulos, D., & Doukidis, G. (2023). A two-stage business analytics approach to perform behavioural and geographic customer segmentation using e-commerce delivery data. *Journal of Decision Systems*, 1-29.
9. Vohra, R., Pahareeya, J., Hussain, A., Ghali, F., & Lui, A. (2020). Using self organizing maps and K means clustering based on RFM model for customer segmentation in the online retail business. In *Intelligent Computing Methodologies: 16th International Conference, ICIC 2020, Bari, Italy, October 2–5, 2020, Proceedings, Part III 16* (pp. 484-497). Springer International Publishing.
10. Casas-Rosal, J. C., Segura, M., & Maroto, C. (2023). Food market segmentation based on consumer preferences using outranking multicriteria approaches. *International Transactions in Operational Research*, 30(3), 1537-1566.

11. Endler, K. D., Scarpin, C. T., Steiner, M. T. A., Sfeir, T. A., & Veiga, C. P. D. (2023). Knowledge discovery in databases: an application to market segmentation in retail supermarkets. *International Journal of Business Intelligence and Data Mining*, 22(3), 310-332.
12. Rogić, S., & Kaščelan, L. (2021). Segmentation Approach for Athleisure and Performance Sport Retailers Based on Data Mining Techniques. *International Journal of E-Services and Mobile Applications (IJESMA)*, 13(3), 71-85.
13. Park, H. J., & Kim, J. G. (2022). An enhanced customer segmentation model for online retailing using RFM analysis and clustering. *Journal of Retailing and Consumer Services*, 67, 102768.'
14. Liu, J., Lu, Y., Chen, Y., & Yu, X. (2021). Customer segmentation based on purchasing behavior analysis using data mining techniques in e-commerce. *Journal of Ambient Intelligence and Humanized Computing*, 12(7), 9937-9950.
15. Luo, Y., Zhong, L., & Zhang, Y. (2021). An improved K-means clustering algorithm for customer segmentation in online retailing. *Journal of Ambient Intelligence and Humanized Computing*, 12(11), 16177-16188.