Neural Computing and Applications
https://doi.org/10.1007/s00521-023-09339-6

ORIGINAL ARTICLE

Customer profiling, segmentation, and sales prediction using AI in direct marketing

Mahmoud SalahEldin Kasem[1] · Mohamed Hamada[2] · Islam Taj-Eddin[3]

Received: 6 March 2023 / Accepted: 26 November 2023

Abstract

In the current business environment, where the customer is the primary focus, effective communication between marketing and senior management is vital for success. Effective customer profiling is a cornerstone of strategic decision-making for digital start-ups seeking sustainable growth and customer satisfaction. This research investigates the clustering of customers based on recency, frequency, and monetary (RFM) analysis and employs validation metrics to derive optimal clusters. The K-means clustering algorithm, coupled with the Elbow method, Silhouette coefficient, and Gap Statistics method, facilitates the identification of distinct customer segments. The study unveils three primary clusters with unique characteristics: new customers (Cluster A), best customers (Cluster B), and intermittent customers (Cluster C). For platform-based Edutech start-ups, Cluster A underscores the importance of tailored learning content and support, Cluster B emphasizes personalized incentives, and Cluster C suggests re-engagement strategies. By understanding and addressing the diverse needs of these clusters, digital start-ups can forge enduring connections, optimize customer engagement, and fuel sustainable business growth.

Keywords Data mining · SVM · Boosting tree · RFM analysis methodology · Deep learning

1 Introduction

In today's business landscape, companies are faced with the challenge of identifying potential customers who are most likely to respond positively to a product or offer, this is where data mining techniques come into play. With the increasing amount of data available, data mining has become an essential tool for direct marketing efforts, allowing companies to create a prediction response model based on past client purchase data. This study aims to present a data mining preprocessing method for developing a customer profiling system that improves the sales performance of an enterprise. The study uses an RFM analysis methodology to evaluate client capital and a boosting tree for prediction. Furthermore, the study highlights the importance of customer segmentation methods and algorithms in increasing the accuracy of the prediction. The main result of this study is the creation of a customer profile and forecast for the sale of goods, which will assist decision-makers in making strategic marketing decisions. The study is expected to provide valuable insights for companies looking to improve their direct marketing efforts and increase sales performance through data mining-based customer profiling [1–3].

The proposed methodology in this study utilizes the RFM analysis (recency, frequency, and monetary) approach to assess client capital, coupled with a boosting tree algorithm for predictive modeling. Additionally, the study emphasizes the crucial role of customer segmentation methods and algorithms in enhancing prediction accuracy. The primary outcome of this research is the development of

✉ Mahmoud SalahEldin Kasem
mahmoud.salah@aun.edu.eg

Mohamed Hamada
M.hamada@iitu.edu.kz

Islam Taj-Eddin
itajeddin@aun.edu.eg

[1] Department of Multimedia Systems, Assiut University, Asyut, Egypt

[2] Department of Computer Science, International IT University, Almaty, Kazakhstan

[3] Department of Information Technology, Assiut University, Asyut, Egypt

Published online: 23 December 2023

a customer profile that offers valuable insights into customer behavior and sales forecasts for goods. However, there is no explicit mention of any secondary results or derivative findings throughout the introduction.

To achieve the research goal of enhancing sales performance through data-driven customer profiling, the study will address a series of key tasks. These tasks include data collection, a comprehensive study of machine learning methods, specifying the structure of client profiles along with their types and relevant indicators, analyzing and organizing customer data, systematizing international practices for improving client profiling, identifying effective methods for researching client profiles, and exploring the concept of "consumer loyalty" in modern marketing. Furthermore, the research aims to clarify the nature and structure of consumer loyalty, highlight foreign experiences in enhancing consumer loyalty, and emphasize the factors influencing the selection of reward systems for developing comprehensive consumer loyalty programs for goods and services manufacturers. Practical recommendations for the formation of such loyalty programs will also be formulated.

Deep learning is a subfield of machine learning that has seen widespread applications in various industries. Deep learning models have been applied to tasks such as text classification, sentiment analysis, machine translation, speech recognition [4], and table detection and recognition [5–8]. Health care is another industry where deep learning has found several applications, including diagnosis, treatment planning, drug discovery [9], and medical imaging analysis [10–12]. In robotics, deep learning is used for autonomous navigation, object recognition [13–15], and robotic control, handwritten recognition for various languages [16–20], questions–answering [21–25], intrusion detection in IoT [26–28], and energy consumption prediction [29, 30].

The research focuses on studying enterprises, organizations, and examining their marketing activities within the context of client policy formation and implementation. Consumers of goods and services are also within the scope of the study. The research delves into the entirety of economic and organizational relationships that arise as firms implement relationship marketing, particularly in creating and implementing programs to build consumer loyalty.

The study's theoretical and methodological foundation is built upon essential research by internal and international scientists in market economy, management, marketing, consumer, and brand loyalty management. Methodologies such as marketing, economic, and statistical analysis, as well as quantitative and qualitative study principles, were utilized. Expert methods were also employed to substantiate the main research provisions. The study also draws inspiration from the work of authors Muller and Hamm [31], emphasizing the significance of starting with segmentation, marketing, and customer data adjustment to achieve accurate analysis and profiling.

The scientific novelty of this research lies in the development of scientific and methodological provisions and recommendations focused on creating and implementing a client profiling framework using AI techniques. Additionally, the study identifies the most effective methods for researching client profiles and enhancing consumer loyalty in Kazakhstani enterprises, exemplifying its applicability in a specific context.

In conclusion, this research aims to provide valuable insights for companies seeking to improve their relationship marketing efforts and enhance sales performance through data-driven customer profiling. With the vast array of customer needs, behavior, and preferences observed in online business platforms, customer segmentation becomes crucial. The study will explore customer segmentation, its importance in understanding customer behavior, and the role of AI in this process, offering practical insights for organizations looking to improve customer retention and benefit upgrades through data-driven customer segmentation.

2 Related work

In the field of customer segmentation, researchers have been experimenting with different algorithms to perform segmentation on customer data. Most of these studies have focused on analyzing customer buying history and purchasing behavior to identify segments. In the following paragraphs, related work methods will be explained followed by a table, i.e., Table 1, that summarizes the advantages and disadvantages.

According to Jiang and Tuzhilin [32], it is crucial to implement both customer segmentation and buyer targeting in order to enhance marketing performance. These two tasks are integrated into a step-by-step approach; however, the challenge of unified optimization arises. To address this issue, the authors proposed the K-classifiers segmentation algorithm. This method prioritizes allocating more resources to those customers who generate the most returns for the company. A significant number of researchers have discussed various techniques for segmenting customers in their studies. Also, the authors propose a direct clustering method for grouping customers. Rather than relying on computed statistics, this approach utilizes transactional data from multiple customers. The authors also acknowledge that finding an optimal segmentation solution is computationally difficult, known as NP-hard. Therefore, Tuzhilin presents alternative sub-optimal clustering methods. The study then experimentally evaluates the customer

Table 1 Related work methods, advantages and disadvantages

Paper	Proposed method	Advantages	Disadvantages
Kashwan [33]	K-means algorithm and a statistical tool	A continuous analysis and online system for e-commerce organization to predict sales	Limited to the use of clustering strategy for determining of market segmentation
Brito [34]	Two data mining methods (clustering and sub-cluster discovery)	Better understanding of customer preferences	Limited to redefined industries
Ballestar [37]	Utilization of cashback and client behavior on social network sites	Shows the reliance on the position of clients inside an organization	Limited to the use of social network writing to promoting like dedication, person-to-person communication, development of client, and commitment of client
Qadadeh [38]	K-means for clustering and self-organized maps for quality of clustering with representation	Involves various procedures for division with expert to further develop organizations	Limited to the use of multiple procedures for segmentation with expert
Christy [40]	RFM analysis and extended to other algorithms such as K-means and RM K-means	Good understanding of the need of client and identification of potential clients for organization	Limited to the use of RFM analysis and extended it to other algorithms such as K-means and RM K-means through minor adjustment in K-means clustering
Jiang [32]	Direct clustering based on transactional data	Identifies customer segments based on actual customer behavior	Finding an optimal segmentation solution is computationally difficult
He [35]	Three-dimensional approach for enhancing CLV, customer satisfaction, and customer behavior	Considers multiple dimensions of customer behavior, leading to more accurate segmentation	Complexity and high computational cost
Sheshasaayee [36]	Integrated approach combining RFM and LTV methods with two-phase approach (statistical and clustering) and neural network	Integrates different methods to improve segmentation	Computationally intensive

segments obtained through direct grouping and finds them to be superior to statistical methods.

Kashwan [33] proposed a K-means algorithm and a statistical tool to propose a model that elaborates on a continuous analysis and online framework for an e-commerce organization to predict sales. They involved a clustering strategy for determining market segmentation because a developed computing-based system is intelligent enough to address results to managers for a quick and fast decision-making cycle.

Brito [34] emphasized that advertising and manufacturing approaches are highly important for customized industries because buying a large variety of products makes it difficult to find specific patterns of customer preferences. As a result, they proposed two different data mining methods, clustering and sub-cluster discovery, for customer segmentation to better understand customer preferences.

He and Li [35] propose a three-dimensional strategy for enhancing customer lifetime value (CLV), customer satisfaction, and customer behavior. The study concludes that consumers have varying needs, and segmentation helps to identify their demands and expectations, which, in turn, leads to providing better service.

Sheshasaayee [36] developed a new integrated approach to segmentation by combining the RFM (recency, frequency, and monetary) and LTV (lifetime value) methods. They employed a two-phase approach, starting with a statistical method in the first phase and then proceeding to cluster in the second phase. The objective is to apply K-means clustering following the two-phase model and then utilize a neural network to improve the segmentation.

Ballestar [37] proposed the role of customers in the use of their cashback and determined the business activity and behavior of customers on the site of a social network. They proposed a model that applied social network analysis to marketing such as loyalty, communication, customer development, and customer engagement to show the dependence of customers' positions within an organization.

Qadadeh [38] proposed the evaluation of data analysis algorithms such as K-means for clustering and self-organized maps for the nature of clustering with visualization. They recommend that involving various procedures for segmentation with experts will further develop organizations such as insurance and study segment elements and behavior of a customer in any customer relationship management dataset.

Several studies have demonstrated the extensive use of RFM technology for customer segmentation and information access. In the context of commercial banks, marketing representatives can employ K-means classification to identify potential customers. To extract valuable insights from customers, data mining methods, including neural networks, C5.0, classification and regression trees, and Chi-squared automatic interaction detector, are highly beneficial for detecting background information related to credit card holders [39].

Christy [40] emphasized that a good understanding of the customer's needs and identification of potential customers for the organization are satisfied by the segmentation process. They performed segmentation using RFM analysis and extended it to other algorithms such as K-means and RM K-means through minor adjustments in K-means clustering.

3 Problem statement

The problem of customer segmentation can be based on various factors such as marketing, sales, support, product, and leadership. Experts in large or small organizations involved in the data analysis process adjust the working group and set the expectations that it will continue to do so in many stages. Some issues that can be resolved through customer segmentation are given below.

- *Marketing* We can solve the problem by understanding our customer base to effectively reach them. We may not be able to observe the business's email lists using the task to be done, but we can observe ones for business to consumer (B2C) subscription organizations with high website traffic volume.
- *Sales* Many issues faced by sales representatives can be resolved by this process. We can route prospects to our self-service stream or the most appropriate group within sales, such as startups, small market businesses, and multi-model businesses, based on clear customer segments.
- *Support* Issues are categorized based on their tool and field. After categorization, it can be used to route support inquiries to the appropriate channels, such as AnswerBot, Alexa, Google Assistant, our help center, or a support representative, to improve customer and business outcomes further.
- *Product* This process can also resolve issues with product quality. Experts should know which product requests and feedback make the biggest impact on which customer and focus accordingly, instead of by volume alone.
- *Leadership* It manages the mission run by e-commerce organizations to deliver their service and make a lead. For this, they create a common language for the product and design and go to markets to describe the customers.

In this paper, we proposed a customer segmentation strategy based on various categories. Different clustering methods such as K-means, repetitive median-based K-means (RM K-Means), and self-organized maps were used for segmentation. We proposed a business model for e-commerce organizations based on segmentation according to various categories and recency, frequency, and monetary (RFM) positioning to retain and acquire customers in e-commerce. Observing new customers is important, but retaining old customers is even more important.

4 Model, tools, environment, and technology

4.1 The customer segmentation approach

Client segmentation is a widely used marketing strategy that involves dividing the customer base into smaller groups based on characteristics such as demographics, behavior, and purchasing history. This enables businesses to understand their customer base better and implement more effective marketing strategies. Vector quantization is an algorithm commonly used for client segmentation, automatically grouping customers based on their behavior data. While it may not always achieve optimal results, it provides valuable insights for businesses to target their marketing efforts. A mapping or vector quantizer can be used to divide data into smaller groups. The mapping is an N-level k-dimensional tool that takes various client RFM values as input vectors. It uses a non-negative real distortion measure to represent the difference between the original and reproduced vectors. The error distortion measure, widely used in mathematical applications, is chosen for its computational efficiency [41].

Data mining techniques have emerged as essential tools in market segmentation. This modern approach to market research involves processing vast datasets from databases using intelligent solutions, such as neural networks, evolutionary algorithms (EA), fuzzy theory, RFM, hierarchical clustering, K-means, bagged clustering, kernel methods, Taguchi method, multidimensional scaling, model-based clustering, and rough sets, among others. These techniques offer highly effective and time-efficient means of segmenting the market [42].

Quantizer optimality is determined based on its ability to minimize average distortion. An N-level quantizer is considered ideal or globally ideal if it achieves this, or at least, it is ideal if, for any remaining quantizers, their distortion is greater than the globally ideal quantizer [40]. Quantizer design aims to obtain an optimal or locally optimal quantizer if possible, and various algorithms have been proposed for this purpose in the literature.

4.2 Machine learning

Interest in machine learning (ML) has grown due to increased processing power and data availability. ML utilizes past experience to enhance performance and make precise predictions. Tasks include classification, regression, ranking, clustering, dimensionality reduction, and complex learning.

- *Data Preprocessing and Model Optimization* It is crucial for AI model creation, involving cleaning, standardization, transformation, feature extraction, and selection.
- *Data Cleaning and Transformation* Class imbalance in ML can lead to issues such as improper evaluation metrics and overfitting. Techniques such as oversampling and undersampling can address this.
- *Missing Data* Handling missing values involves deletion or imputation with estimated values.
- *Sampling* Preprocessing plays a vital role in AI model creation, impacting performance and interpretability. Techniques such as oversampling and undersampling can address class imbalance but have drawbacks.
- *Feature and Variable Selection* Feature selection is critical for identifying relevant data, improving predictive performance, and efficiency. Various methods can be used based on the dataset and computational resources.

4.3 Model for customer segmentation and market segmentation

Commonly used models for customer segmentation include:

- Demographic segmentation;
- Recency, frequency, and monetary (RFM) segmentation;
- Customer status and behavioral segmentation.

Segmentation based on gender is a simple yet effective way for organizations to categorize their customer base, allowing for targeted content and promotions for gender-specific events. RFM segmentation is widely used in the direct mail industry to rank customers based on their purchasing history, considering recency, frequency, and monetary value of purchases [34].

Client status and behavior analysis involve categorizing clients into active and lapsed based on their last purchase. Behavioral analysis analyzes past customer behavior, such as shopping habits and brand preferences, to predict future actions. Data analysts use this information to segment clients into categories and develop strategies for customer retention and acquisition [43].

Market division, first defined in 1956, is a method used by organizations to categorize customers based on similar characteristics, such as geographic location, demographics, product usage, and purchasing behavior. The goal is to increase customer satisfaction and maximize efficiency by tailoring marketing efforts to specific segments. One common tool used in the market division is clustering, which groups elements with similar values into segments [44–46]. While early market division studies only considered one set of factors, modern market division models take into account multiple sets of factors simultaneously, called cooperative market division. There are various market division methods, including K-means clustering, hierarchical clustering, association rule mining, decision trees, and neural networks. The objective is to identify and describe customer groups and reach profitable customer segments [47].

Clustering data is a significant aspect of data mining techniques, involving the utilization of latent class analysis (LCA), prior clustering, and various similarity or distance measures to segment large customer groups based on individual expectations [48]. Sánchez-Fernández [49] presents a conceptual framework centered around tourists' perception of sustainability policies at various destinations, along with a multidimensional measure to assess this construct. Through an empirical analysis conducted across five Mediterranean destinations, the proposed conceptual model was validated, offering substantial empirical evidence supporting the viability of perceived sustainability as a valuable factor in segmentation studies.

4.4 Customer segmentation and client profiling

Market and customer segmentation are often used interchangeably, with market segmentation seen as a high-level strategy and customer segmentation providing a more granular view. The RFM model is a valuable tool for combining customer segmentation and targeting in campaigns [34]. Genetic algorithms can enhance the customer division and targeting process, using the LTV model as a fitness function [38]. Various mathematical methods have been explored for customer segmentation, including statistical techniques, neural networks, genetic algorithms, and K-means fuzzy clustering.

Client profiling involves analyzing client characteristics for tailored marketing strategies, contributing to customer

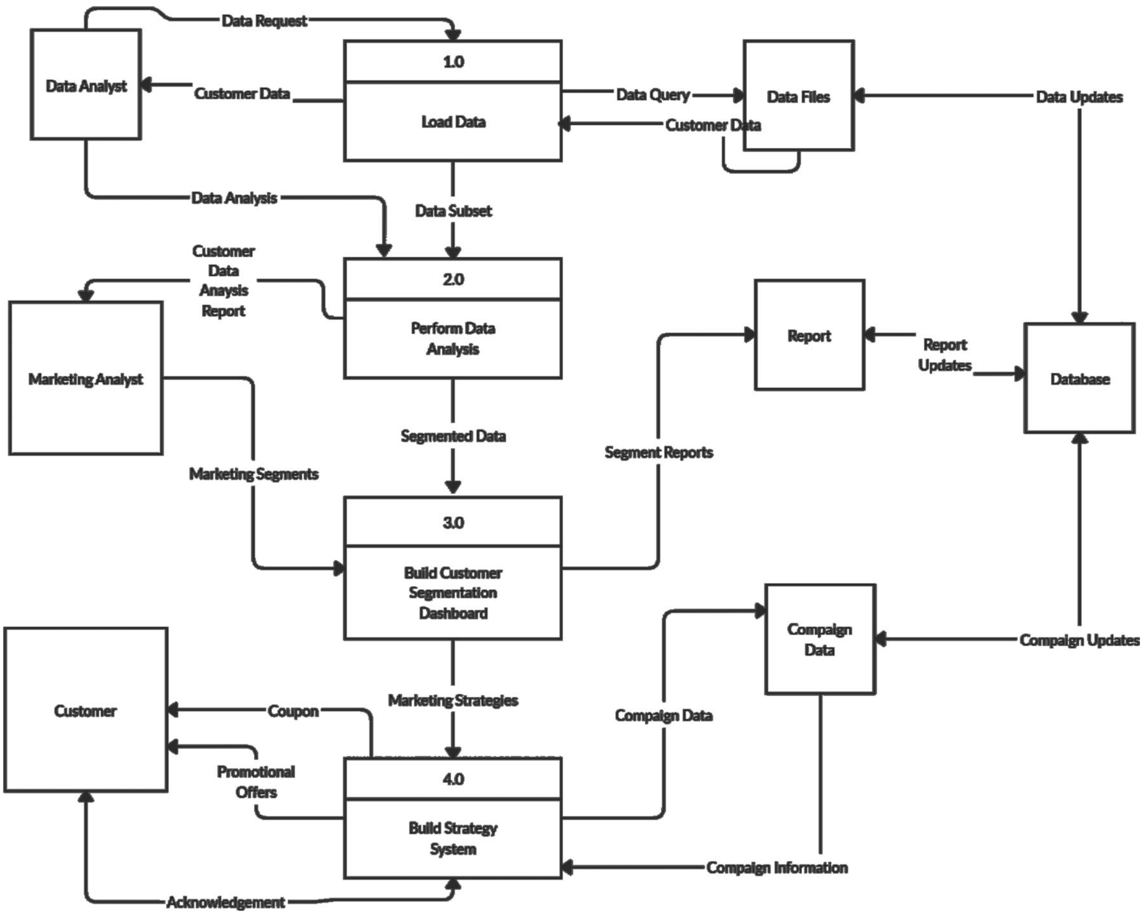


Fig. 1 Processes of customer segmentation

retention and CRM [50]. Segment profiling focuses on understanding specific customer group attributes to guide marketing strategies. Buyer behavior profiles consider social factors such as timing, benefits sought, usage rate, loyalty, and attitude for targeted marketing efforts [51].

The proposed architecture of customer profiling, depicted in Fig. 1, offering a comprehensive approach to understanding customer behavior and preferences.

5 Results and discussion

5.1 Dataset

The data utilized for our research were sourced from the Marketing Campaign dataset[1], which encompasses a cross-border dataset that encompasses several key demographic attributes, including age, education level, ID, annual income, marital status, and presence of children in the household, as shown in Table 2.

The RFM model employed in this study utilized data from the SAS Institute to calculate the recency, frequency, and monetary rankings, enabling the segmentation of customers into distinct groups. The data comprise the following attributes as shown in Table 3.

5.2 Preprocessing

This study employs three distinct algorithms to perform customer clustering utilizing RFM analysis. Initially, the data undergo preprocessing to eliminate outliers and extract pertinent instances. Outliers are detected using the z-score method, which assesses data’s proximity to its mean and standard deviation. This relationship is transformed into a scale from 0 to 1, with values deviating significantly from the mean (zero) identified as outliers.

5.3 Customer profiling approach

Within this investigation, following the completion of data preprocessing, the dataset proceeds to the customer profiling phase. In this phase, the K-means algorithm is

[1] https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign?select=marketing_campaign.csv.

Table 2 Attributes of first datasets

Serial no.	Attributes
1	ID
2	Year_Birth
3	Education
4	Marital_Status
5	Income
6	Kidhome
7	Teenhome

Table 3 Attributes of second datasets

Serial no.	Attributes
1	Dt_Customer
2	Recency
3	MntWines
4	MntFruits
5	MntMeatProducts
6	MntMeatProducts
7	MntFishProducts
8	MntSweetProducts
9	MntGoldProds
10	NumDealsPurchases
11	NumWebPurchases
12	NumCatalogPurchases
13	NumStorePurchases
14	NumWebVisitsMonth
15	AcceptedCmp1
16	AcceptedCmp2
17	AcceptedCmp3
18	AcceptedCmp4
19	AcceptedCmp5
20	Complain
21	Z_CostContact
22	Z_Revenue
23	Response

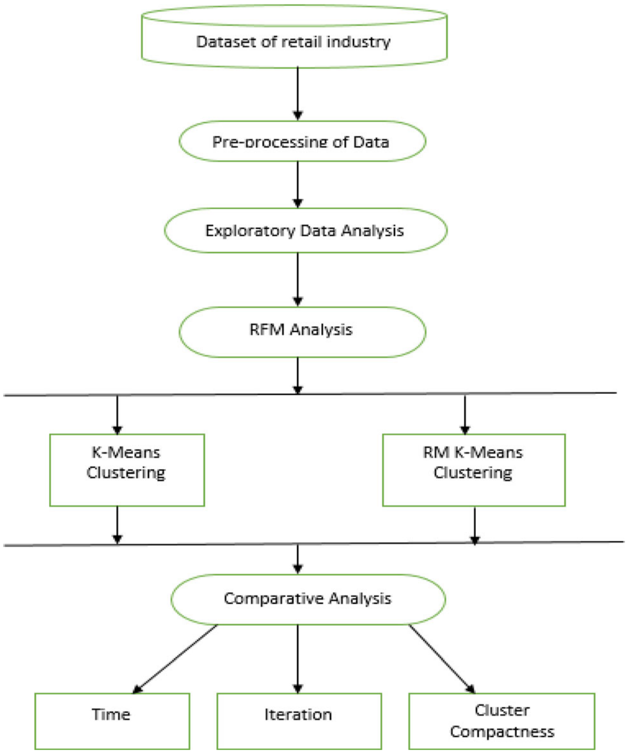


Fig. 2 Processes of RFM analysis

applied to the dataset for clustering purposes, process of K-means on RFM analysis shown in Fig. 2. Subsequently, the outcomes of the K-means clustering are subjected to validation procedures aimed at determining the optimal cluster value (K). This validation is executed using three distinct metrics: the Elbow method, the Silhouette coefficient, and the Gap Statistics method. The graphical representation of these validation results is depicted in Fig. 3 for the Elbow method, Fig. 4 for the Silhouette coefficient, and Fig. 5 for the Gap Statistics method. Importantly, the Matthews correlation coefficient (MCC) scorer, known for its ability to accommodate classes of varying sizes, is employed to assess the effectiveness of the chosen methodology.

The outcome of the Elbow method analysis illustrates a decline in the value of cluster inertia, also known as the sum of squared errors (SSE), as the number of clusters increases. From the graphical representation, it becomes apparent that potential candidates for the optimal K value reside within the range of $K = 2$ and $K = 8$. This observation is guided by the appearance of a discernible "elbow" shape in the graph, where the decrease in SSE starts to plateau. Nonetheless, the validation process remains essential and will involve the assessment of the other two metrics for confirming the optimal cluster configuration.

The second validation involves employing the Silhouette coefficient method, which yields the Silhouette scores for each cluster as presented in Table 4. This method allows for a comparative assessment with DP Agustino [52].

The Matthews correlation coefficient (MCC) for the test data was computed as 0.88, suggesting potential challenges in accurately categorizing positive examples within the test set.

Based on the outcomes of the clustering analysis, the findings reveal the existence of three distinct clusters that serve as foundational categories for digital start-ups in executing customer profiling strategies. The initial category (designated as Cluster A) pertains to new customers,

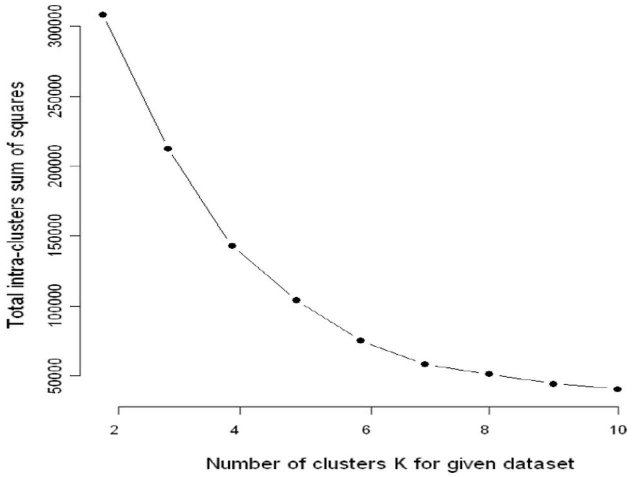


Fig. 3 Elbow method

denoting those who have initiated their first purchase of products within the digital start-up's domain. In the context of this research, it is imperative to augment engagement with these new customer segments by aligning strategies with their specific needs, thereby enhancing the prospect of subsequent product repurchases. Particularly, for platform-based Edutech start-ups, augmenting the platform with enriched educational content and providing tailored teacher support emerge as a pivotal strategy to amplify customer relevance and sustained interest.

The second category (identified as Cluster B) encompasses the best customers. This category comprises individuals who have engaged in multiple purchases, particularly emphasizing recent transactions. Customers within this cluster exhibit a pronounced potential to be offered each new product launch by the digital start-up. Furthermore, to foster more robust customer loyalty, the provision of exclusive discounts to customers within this category presents an effective approach.

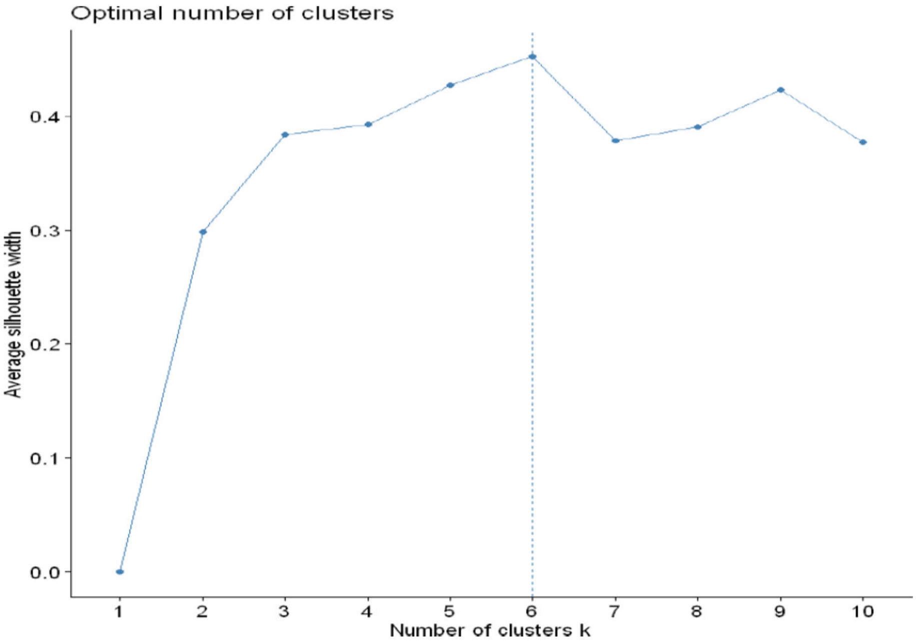
Cluster C, the third category, is composed of intermittent customers. These customers display a sporadic pattern of engagement, characterized by occasional purchases and fluctuations in their interaction frequency. For digital start-ups, devising targeted marketing efforts that encourage consistent engagement from these intermittent customers can enhance their loyalty and transform them into more regular purchasers. Tailored promotions and personalized offers are instrumental in motivating these customers to establish a more enduring connection with the digital start-up's offerings.

6 Conclusion

In conclusion, this research delved into the critical domain of customer profiling within the context of digital start-ups. Through a comprehensive analysis of clustering algorithms and validation methodologies, we successfully identified distinct customer clusters that offer invaluable insights for tailored business strategies. The utilization of K-means clustering, coupled with validation metrics such as the Elbow method, Silhouette coefficient, and Gap Statistics method, provided a robust foundation for customer segmentation.

The results revealed three primary clusters that serve as significant touchpoints for digital start-ups to refine their customer engagement tactics. Cluster A, representing new

Fig. 4 Silhouette method



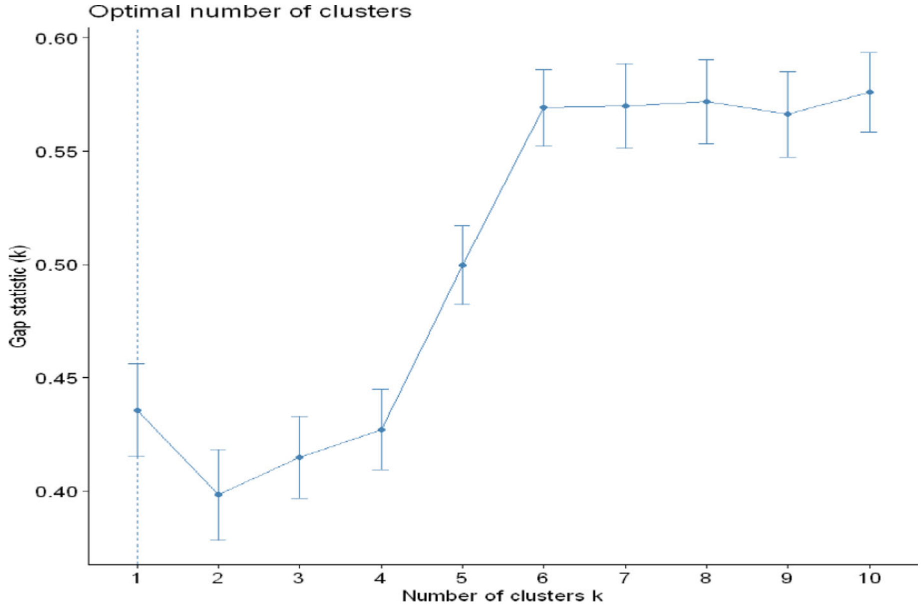


Fig. 5 Gap Statistics method

Table 4 Silhouette scores for different *K* values

K value	Silhouette score for Agustino [52]	Silhouette score for our model
2	0.87	.88
3	0.52	0.55
4	0.22	0.22
5	0.56	0.55
6	0.62	0.64
7	0.70	0.71
8	0.70	0.69
9	0.71	0.72
10	0.65	0.66

customers, necessitates tailored approaches to enhance their initial experience and foster repurchase potential. In the realm of platform-based Edutech start-ups, offering enriched learning content and personalized support emerges as a potent strategy. Cluster B, housing the best customers, signifies a vital avenue for product promotion and customer loyalty enhancement. Customized incentives and exclusive offerings can solidify their engagement and elevate their lifetime value. Cluster C, comprising intermittent customers, highlights an opportunity to re-engage and cultivate consistency. Strategic interventions, such as targeted promotions and individualized incentives, can transform intermittent customers into steady patrons.

In the broader landscape of digital start-ups, the outcomes underscore the paramount importance of customer profiling in enhancing business outcomes. By acknowledging the nuanced requirements of different customer clusters, start-ups can forge more meaningful and enduring connections, thereby fostering growth, customer satisfaction, and long-term success. As the digital landscape continues to evolve, these insights hold the potential to guide start-ups toward informed decisions that resonate with their customer base, fostering a symbiotic relationship between innovation and consumer needs.

7 Future work

In the future work, more advanced methods for predicting customer churn may be explored, such as weighted random forests and hybrid models that can handle unstructured data. This would enable the extraction of relevant attributes for potential customer segmentation studies in the retail industry. As highlighted in the literature review, using hybrid models has shown promising performance gains and could be a strategy to improve the models.

Artificial intelligence has the potential to revolutionize various industries by transforming existing business processes and creating new business models. Key areas of focus include consumer engagement, digital manufacturing, smart cities, autonomous vehicles, risk management, computer vision, and speech recognition. AI has already demonstrated positive results in a range of sectors including health care, law enforcement, finance, security, trade, manufacturing, education, mining, and logistics.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no conflicts of interest related to this work.

References

1. Alsayat A (2023) Customer decision-making analysis based on big social data using machine learning: a case study of hotels in mecca. Neural Comput Appl 35:4701–4722
2. Kalkan IE, Şahin C (2023) Evaluating cross-selling opportunities with recurrent neural networks on retail marketing. Neural Comput Appl 35(8):6247–6263
3. Das S, Nayak J (2021) Customer segmentation via data mining techniques: state-of-the-art review. Comput Intell Data Min: Proc ICCIDM 2022:489–507
4. Chorowski J, Bahdanau D, Serdyuk D, Cho K, Bengio Y (2015) Attention-based models for speech recognition, arXiv preprint arXiv:1506.07503
5. Abdallah A, Berendeyev A, Nuradin I, Nurseitov D (2022) Tncr: table net detection and classification dataset. Neurocomputing 473:79–97
6. Prasad D, Gadpal A, Kapadni K, Visave M, Sultanpure K (2020) Cascadetabnet: an approach for end to end table detection and structure recognition from image-based documents. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp 572–573
7. Kasem M, Abdallah A, Berendeyev A, Elkady E, Abdalla M, Mahmoud M, Hamada M, Nurseitov D, Taj-Eddin I (2022) Deep learning for table detection and structure recognition: a survey, arXiv preprint arXiv:2211.08469
8. Abdimanap G, Bostanbekov K, Abdallah A, Alimova A, Kurmangaliyev D, Nurseitov D (2022) Enhancing core image classification using generative adversarial networks (gans), arXiv e-prints arXiv–2204
9. Fakoor R, Ladhak F, Nazi A, Huber M (2013) Using deep learning to enhance cancer diagnosis and classification. In: Proceedings of the international conference on machine learning, volume 28, ACM, New York, USA, pp 3937–3949
10. Nie L, Wang M, Zhang L, Yan S, Zhang B, Chua T-S (2015) Disease inference from health-related questions via sparse deep learning. IEEE Trans Knowl Data Eng 27:2107–2119
11. Abdallah A, Kasem M, Hamada MA, Sdeek S (2020) Automated question-answer medical model based on deep learning technology. In: Proceedings of the 6th International Conference on Engineering & MIS 2020, pp 1–8
12. Yu L, Hermann KM, Blunsom P, Pulman S (2014) Deep learning for answer sentence selection, arXiv preprint arXiv:1412.1632
13. Logothetis NK, Sheinberg DL (1996) Visual object recognition. Annu Rev Neurosci 19:577–621
14. Nurseitov D, Bostanbekov K, Abdimanap G, Abdallah A, Alimova A, Kurmangaliyev D (2022) Application of machine learning methods to detect and classify core images using gan and texture recognition, arXiv preprint arXiv:2204.14224
15. Mahmoud M, Kang H-S (2023) Ganmasker: a two-stage generative adversarial network for high-quality face mask removal. Sensors 23:7094
16. Mahmoud SA, Ahmad I, Al-Khatib WG, Alshayeb M, Parvez MT, Märgner V, Fink GA (2014) Khatt: an open arabic offline handwritten text database. Pattern Recogn 47:1096–1112
17. Nurseitov D, Bostanbekov K, Kurmankhojayev D, Alimova A, Abdallah A, Tolegenov R (2021) Handwritten Kazakh and Russian (hkr) database for text recognition. Multimed Tools Appl 80:33075–33097
18. Toiganbayeva N, Kasem M, Abdimanap G, Bostanbekov K, Abdallah A, Alimova A, Nurseitov D (2022) Kohtd: Kazakh offline handwritten text dataset. Sig Process Image Commun 108:116827
19. Abdallah A, Hamada M, Nurseitov D (2020) Attention-based fully gated cnn-bgru for Russian handwritten text. J Imag 6:141
20. Daniyar Nurseitov GA, Kairat B, Maksat K, Anel A, Abdelrahman A (2020) Classification of handwritten names of cities and handwritten text recognition using various deep learning models. Adv Sci Technol Eng Syst J 5:934–943
21. Karpukhin V, Oğuz B, Min S, Lewis P, Wu L, Edunov S, Chen D, Yih W-t (2020) Dense passage retrieval for open-domain question answering, arXiv preprint arXiv:2004.04906
22. Chen D, Yih W-t (2020) Open-domain question answering. In: Proceedings of the 58th annual meeting of the association for computational linguistics: tutorial abstracts, pp 34–37
23. Abdallah A, Jatowt A (2023) Generator-retriever-generator: A novel approach to open-domain question answering, arXiv preprint arXiv:2307.11278
24. Abdallah A, Abdalla M, Elkasaby M, Elbendary Y, Jatowt A (2023a) Amurd: annotated multilingual receipts dataset for cross-lingual key information extraction and classification, arXiv preprint arXiv:2309.09800
25. Abdallah A, Piryani B, Jatowt A (2023) Exploring the state of the art in legal qa systems, arXiv preprint arXiv:2304.06623
26. Mahmoud M, Kasem M, Abdallah A, Kang HS (2022) Ae-lstm: autoencoder with lstm-based intrusion detection in iot, in, (2022) International Telecommunications Conference (ITC-Egypt). IEEE, pp 1–6
27. Xu W, Jang-Jaccard J, Singh A, Wei Y, Sabrina F (2021) Improving performance of autoencoder-based network anomaly detection on nsl-kdd dataset. IEEE Access 9:140136–140146
28. Akkad A, Wills G, Rezazadeh A (2023) An information security model for an iot-enabled smart grid in the saudi energy sector. Comput Electr Eng 105:108491
29. Waschneck B, Reichstaller A, Belzner L, Altenmüller T, Bauernhansl T, Knapp A, Kyek A (2018) Optimization of global production scheduling with deep reinforcement learning. Proc Cirp 72:1264–1269
30. Hamada MA, Abdallah A, Kasem M, Abokhalil M (2021) Neural network estimation model to optimize timing and schedule of software projects. In: 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), IEEE, pp 1–7

31. Müller H, Hamm U (2014) Stability of market segmentation with cluster analysis-a methodological approach. Food Qual Prefer 34:70–78
32. Jiang T, Tuzhilin A (2008) Improving personalization solutions through optimal segmentation of customer bases. IEEE Trans Knowl Data Eng 21:305–320
33. Kashwan KR, Velu C (2013) Customer segmentation using clustering and data mining techniques. Int J Comput Theory Eng 5:856
34. Brito PQ, Soares C, Almeida S, Monte A, Byvoet M (2015) Customer segmentation in a large database of an online customized fashion business. Robot Comput-Integr Manuf 36:93–100
35. He X, Li C (2016) The research and application of customer segmentation on e-commerce websites. In: 2016 6th International Conference on Digital Home (ICDH), IEEE, pp 203–208
36. Sheshasaayee A, Logeshwari L (2017) An efficiency analysis on the tpa clustering methods for intelligent customer segmentation. In: 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), IEEE, pp 784–788
37. Ballestar MT, Grau-Carles P, Sainz J (2018) Customer segmentation in e-commerce: applications to the cashback business model. J Bus Res 88:407–414
38. Qadadeh W, Abdallah S (2018) Customers segmentation in the insurance company (tic) dataset. Proc Comput Sci 144:277–290
39. Lu Z, Peiyi W, Ping C, Xianglong L, Baoqun Z, Longfei M (2019) Customer segmentation algorithm based on data mining for electric vehicles. In: 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), IEEE, pp 77–83
40. Christy AJ, Umamakeswari A, Priyatharsini L, Neyaa A (2021) Rfm ranking-an effective approach to customer segmentation. J King Saud Univ-Comput Inform Sci 33:1251–1257
41. Pranata I, Skinner G (2015) Segmenting and targeting customers through clusters selection & analysis. In: 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE, pp 303–308
42. Dutta S, Bhattacharya S, Guin KK (2015) Data mining in market segmentation: a literature review and suggestions. In: Proceedings of Fourth International Conference on Soft Computing for Problem Solving: SocProS 2014, Volume 1, Springer, pp 87–98
43. Tsao Y-C, Raj PVRP, Yu V (2019) Product substitution in different weights and brands considering customer segmentation and panic buying behavior. Ind Mark Manage 77:209–220
44. Liu Y, Kiang M, Brusco M (2012) A unified framework for market segmentation and its applications. Expert Syst Appl 39:10292–10302
45. Kim S-Y, Jung T-S, Suh E-H, Hwang H-S (2006) Customer segmentation and strategy development based on customer lifetime value: a case study. Expert Syst Appl 31:101–107
46. Weinstein A (2013) Handbook of market segmentation: Strategic targeting for business and technology firms, Routledge
47. Hosseini M, Shabani M (2015) New approach to customer segmentation based on changes in customer value. J Market Anal 3:110–121
48. Swenson ER, Bastian ND, Nembhard HB (2018) Healthcare market segmentation and data mining: a systematic review. Health Mark Q 35:186–208
49. Sánchez-Fernández R, Iniesta-Bonillo MÁ, Cervera-Taulet A (2019) Exploring the concept of perceived sustainability at tourist destinations: a market segmentation approach. J Travel Tour Market 36:176–190
50. Romdhane LB, Fadhel N, Ayeb B (2010) An efficient approach for building customer profiles from business data. Expert Syst Appl 37:1573–1585
51. Tong L, Wang Y, Wen F, Li X (2017) The research of customer loyalty improvement in telecom industry based on nps data mining, China. Communications 14:260–268
52. Agustino DP, Harsemadi IG, Budaya IGBA (2022) Edutech digital start-up customer profiling based on rfm data model using k-means clustering. J Inform Syst Inform 4:724–736

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.