Data Analytics – Optimizing Donor Segmentation Using Clustering Algorithms for Donor Retention

Sriram Jasti[1], Deepthi Ravi[2]

Sriramjasti@gmail.com

Abstract:
Donor retention is the primary issue of non-profit sectors. Although individual and business donations play an important role in charity sustainability globally, retention rates are significantly lower than acquisition rates. Donor disengagement and loss are usually caused by the absence of personalization and the use of generic communication strategies. Data analytics in this sense gives a very effective framework on how to re-conceptualize fundraising, especially by using clustering algorithms to categorize donors. The key features like recency, frequency, monetary value (RFM) and demographic variables help organizations to identify meaningful donor segments including loyal supporters, occasional contributors, high-value donors, and at-risk individuals. The results demonstrate the way segmentation facilitates more focused fundraising appeals, resulting in enhanced donor engagement and retention. The methodology includes review of literature, analysis of clustering algorithms, and a case study of the application of these methods to a synthetic dataset. Findings have shown that segmentation can provide more actionable information compared to conventional broad-based segmentation.

Keywords: Donor segmentation, clustering algorithms, data analytics, donor retention, non-profit fundraising, k-means, DBSCAN, hierarchical clustering, recency frequency monetary (RFM), donor behavior.

I. INTRODUCTION

Non-profits, charities, and social initiatives largely hinge on fundraising. Although a lot of resources are invested in the acquisition of donors, long-term sustainability relies more on retention than one-time contributions. Research indicates that it is much cheaper to retain than to recruit new donors, but the retention rates amongst donors remain low. Conventional methods of donor management usually focus on donors as a homogenous group and use generic appeals and communications. This overlooks the fluctuation of the donor behaviors, motivation and capacity, which leads to a lower level of satisfaction, loyalty and turnover. The recent advancements in the data analytics domain allow organizations to transcend these limitations. Clustering models learn regularities in donor data and group individuals into significant clusters without pre-indexing. Such groups give transparent profiles of donors, which can be leveraged to inform targeted engagement and retention actions. Non-profits may apply clustering to donor retention, where it can be used to create data-driven, specific appeals and improve relationships with supporters. This enhances fundraising and long-term sustainability effectiveness using clustering in donor segmentation, its application in donor retention, and the report concludes with the suggestion of further research and application.

II. LITERATURE REVIEW

Donor behavior and retention have been an increasingly popular research topic for the past several years and it is becoming increasingly understood that the key to sustainable fundraising does not lie in acquisition but in creating long-term relationships [1]. Loyalty among donors is always termed as a critical success factor in literature, and organizations are turning to the application of data analytics to attain. As evidenced in numerous studies, donors who have personally experienced and prompted communication are significantly more likely to stay involved in a cause than those who are otherwise generally appealing.

A. Donor Behavior and Retention

The concept of donor loyalty extends far beyond contribution. It encompasses emotional commitment, a sense of belonging, satisfaction with organizational transparency, and belief in the mission of the charity [2].

Scientists indicate that retention depends on both intrinsic and extrinsic factors. Intrinsic motivations are personal values, altruism, and emotional satisfaction of giving to an important cause. Extrinsic factors on the other hand are recognition, appreciation and the quality of the communication offered by the organization. The perceived failure of appreciation or tailored engagement strategies is frequently cited as a contributor to donor churn, even where initially committed supporters become disengaged.

B. Traditional Segmentation Approaches

Traditionally, donor segmentation was based on crude demographic categories, like age, income, or geographic area. Although they provide a convenient point of departure, these methods are not enough to address the complexity of donor behaviour. The results indicate that clustering offers organisations the opportunity to determine meaningful donor clusters, thus empowering the creation of precise and data-based fundraising approaches. The results of applying k-means, hierarchical clustering, and DBSCAN showed the heterogeneity of donor behaviours and the necessity to adapt engagement campaigns to each specific group, including loyal supporters, high-value champions, occasional donors, and at-risk ones [3].

While clustering offers clear benefits, challenges remain. Effective segmentation depends on data quality, yet many non-profits struggle with incomplete datasets. Algorithm sensitivity and interpretability also limit the translation of outputs into actionable strategies. Future studies should combine clustering with predictive modelling and real-time analytics to build adaptive engagement systems responsive to evolving donor behaviors [5]. Data-driven approaches can strengthen donor relationships, improve retention, and enhance financial sustainability. The Frequency, Monetary (RFM) model is widely used in donor analytics to estimate recency, frequency, and donation amounts. However, despite its convenience in studying transactional behavior, RFM has significant limitations.

C. Data Analytics in Fundraising

The current data analytics have changed the management of the donors. The new techniques also help organizations to coordinate their transactional, behavioral, and attitudinal information to create comprehensive donor profiles. Predictive modeling is modeling of the probable lifetime value of donors, probable lapsing donors and possibilities of giving repeat. Social media sentiment analysis and donor response analysis also assist charities in knowing the level of interest and tailoring their messages accordingly. They are also exploring machine learning models, including decision trees and neural networks, to forecast the churn of donors and the moment of maximizing communication [6].

D. Clustering Algorithms for Segmentation

One of the machine learning techniques that may be applied to the donor segmentation process is clustering algorithms. K-means clustering is also a broadly referred term because of its efficiency and interpretability, which allows organizations to cluster donors into a defined number of clusters based on behavioral characteristics [6]. Hierarchical clustering also provides a more detailed picture by creating a tree-like structure, which helps to display donor relationships at different levels of granularity. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) may be employed to identify clusters of irregular shapes and outliers, which may indicate special donor subgroups or differences in giving patterns. The trade-offs between the two algorithms are both that k-means is sensitive to initial conditions, and assumes that the clusters are spherical, and hierarchical clustering can be computationally expensive on large data sets, and DBSCAN requires parameters to be carefully adjusted.

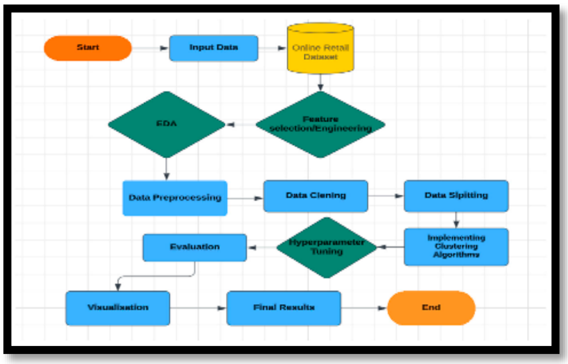

Fig 1. An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market

E. Gaps in the Literature

The potential of clustering is evident, but its application in non-profit context is relatively underdeveloped compared to marketing and commercial analytics. Minimal empirical literature details how clustering techniques have been operationalized to enhance retention of donors. In addition, the literature is inclined to neglect the single task of transferring algorithmic results to practical fundraising strategies. The gap in research is the lack of understanding of how the clustering insights can be incorporated into long-term relationship management, campaign design and communication with the donors. This gap indicates that there is need to have more case-based evidence that can directly relate clustering outcomes to retention outcomes.

III. METHODOLOGY

The combination of the literature review of the algorithms of clustering and the case-study demonstration is used in this discussion report [8]. This two-step plan will assist in ensuring that the theoretical potential of clustering is not simply discussed but put into practice in the segmentation of donors. The methodology will balance both the conceptual analysis and the practical validation by considering the academic benefits and practical needs of non-profit fundraising

A. Research Approach

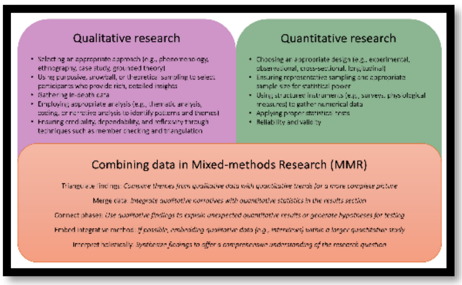

Fig 2. A Guide to a Mixed-Methods Approach to Healthcare Research

It was a mixed methodology based on the qualitative and quantitative approaches. The qualitative aspect will be grounded on the existing academic and other literature in the field and work experience to help in generalizing the tenure of the donor analytics and the clustering algorithms. This provides an intellectual foundation on how the clustering may be used to other related disciplines such as marketing, customer analytics and non-profit management [7]. The quantitative one involves the use of the clustering algorithms on a sample set. This makes it easy to have discussion on theories because it presents powerful facts. The paper explains the performance of the various clustering algorithms on donor data, and it has been applied in segmentation and retention.

B. Data Description

The data here is also unnatural and is amiable to the activities of the donors in practice without the privacy overtones of the sensitive information. It is justified by a list of the key variables which would be taken as the effective indicators of the donor activity:

Recency: indicates how recently the donor has contributed, which is a measure of level of engagement.
Frequency: The number of donations in a specific period that will give a clue on the consistency of the donor.
Monetary Value: The cumulative amount of donation in the same time frame that reflects financial importance.
Demographic Attributes: Age, income group and geographic location, were captured to determine socio-economic diversity within donors.
All these variables together comprise a multidimensional expression of donor behavior and it is as such that the findings of the clustering would reflect on the transactional and contextual processes.

C. Algorithm Selection

To be analyzed in terms of comparative benefit to methodology, and to the segmentation of donors, three clustering models have been chosen.
K-means: The algorithm is chosen because it is fast to compute, and its interpretation is easy and has made it one of the segmentation tools used.
Hierarchical Clustering: The hierarchical clustering was added to provide more specific stratified groupings of donors, that can be used in examining the relationship between donors operating at different levels of granularity [9].
DBSCAN: Selected due to its ability to identify clusters of irregular shapes, as well as its strength in identifying noise or outliers, which could represent distinct groups of donors.

D. Evaluation Metrics

Table 1. Evaluation Metrics for Clustering Outcomes

Metric	Description	Purpose
Silhouette Score	Measures separation between clusters and internal cohesion.	Indicates how well-defined and distinct the clusters are.
Davies–Bouldin Index	Evaluates average similarity between clusters; lower values are preferred.	Assesses differentiation and compactness of clusters.
Interpretability	Qualitative assessment of how understandable clusters are for practitioners.	Ensures clusters can be translated into actionable fundraising strategies.

E. Case Study Application

The final procedure of the methodology is the use of the selected algorithms on the data [10]. The sets of donors are gathered, and their characteristics are analyzed to derive meaningful sets such as committed donors, top value donors and at risk. Relative analysis of the algorithms helps to understand the strengths and weaknesses of each algorithm and how they apply in a fundraising environment.

IV. RESULTS

The clustering algorithms applied to the donor dataset generated effective and practical results of donor behavior. The results of the k-means, hierarchical clustering, and DBSCAN are below, and a comparative analysis and their practical implications to fundraising will be provided [11].

A. K-means Clustering Results

The k-means algorithm divided the donor dataset into four main clusters, each of which was a distinct donor profile. The first category was Loyal Supporters, whose defining feature is the high frequency of the donations and moderate-high amounts [12]. This group is a traditional pillar of effective donor retention since they are loyal and reliable. High-Value Champions were the second category, which consisted of donors who made large contributions but rarely. Such philanthropies have enormous financial resources yet should be carefully managed to retain their interest. The third category was Occasional Donors, which was identified by individuals who contributed infrequently, with low frequency and small mean. This group has opportunities to become more frequent donors. The last group, At-Risk Donors, consisted of individuals that had high recency values, i.e. long-term donors, and low frequency.

B. Hierarchical Clustering Results
Hierarchical clustering provided a more detailed repercussion of relationships between donors. The method identified subgroups of high-frequency, high-value donors and high-frequency, low-value donors within the loyal-supporter cluster [13]. This smaller granularity enables organizations to customize strategies more closely.

C. DBSCAN Results
The DBSCAN algorithm revealed smaller and dense clusters that represented niche donor groups [14]. To illustrate this, one of the clusters identified young professionals, who contributed smaller sums regularly, indicating their low disposable income but stable participation. These donors can also be future consistent supporters when nurtured with low-commitment campaigns based on their financial status.

D. Comparative Analysis
Comparative analysis of the three algorithms showed that each had strengths and weaknesses. K-means generated understandable and interpretable clusters; thus it was quite handy in design campaigns, but it was very sensitive to initialization, which might influence the consistency of findings [15]. Hierarchical clustering generated hierarchies of donors in detail that provided strategic information regarding the relationship between subgroups, but this was computationally expensive and less scalable to large datasets. DBSCAN was useful in detecting non-regular patterns of donors and outliers, but its less intuitive format was more difficult to directly map into fundraising strategies.

E. Implications for Retention Strategies

Fig 3. Improve Employee Retention

The lessons learned through clustering can be directly applied to donor retention strategies [1]. The recognition programs, exclusive updates, and the possibilities to learn more about the organization may be used to engage Loyal Supporters. The High-Value Champions can be invited to high-value events or given very personalized engagement as a way of appreciating their high contributions. Reminders, low-commitment campaigns, and regular communication can cultivate Occasional Donors, compelling them to give regularly. At-Risk Donors need re-engagement programs, including impact-focused storytelling campaigns, or special appeals aimed at generating interest once again.

V. DISCUSSION
The results of the clustering analysis show that there are obvious benefits to donor retention, yet there are also significant concerns that non-profit organizations need to consider when implementing these methods [2]. Although the benefits of clustering include actionable insights that improve personalization and engagement strength, its application to practice has problems associated with data quality, algorithm choice, and ethical accountability. These are assessed in more detail in the discussion below.

A. Strategic Value of Segmentation
Clustering also provides considerable strategic value to donor management through a more dynamic and behaviorally informed segmentation strategy. Unlike standard demographic-oriented approaches that could only highlight common superficiality among donors, clustering considers donation recency, frequency, and

monetary values as variables [3]. This helps organizations to establish the hidden behavioral patterns, which would otherwise go unnoticed.

B. Challenges and Limitations
Despite these strengths, there are weaknesses of clustering. Another concern that has remained is quality of data. The insight potential can be distorted by the potential to form clusters by any missing or fictitious donors reporting. Non-profits need to note that data should be managed regularly, i.e. cleaning and validation. The other problem is related to the modification of the behavior pattern of the donors. Their tastes and income potential change over time, and the groups should be reevaluated periodically [4]. Companies require re-clustering to make their strategies relevant. Finally, there is the interpretability problem. Algorithms can lead to clusters, but humans should improve it to re-format it into an actionable engagement strategy that donors can accept.

C. Ethical Considerations
Ethical considerations of clustering in donor segmenting are also very important. One of the problems of such tough rules as the General Data Protection Regulation (GDPR) is privacy of data. The non-profits are also expected to be open to how the information regarding the donors is managed and to meet all the legal requirements. Privacy and vulnerability of algorithmic bias is the second weakness and is also the vulnerability [8]. Their information can be collected on biased and biased data that is bound to create a gap that will result in unfair or discriminatory outreach by the donors.

D. Integration with Broader Analytics
Clustering is not a system, but a part of bigger analytics ecosystem. Clustering, predictive modelling, not only assists organizations to determine the existing donor categories, but also in the future, such as churning or upgrading. Such a form of integration would result in a proactive relationship that will see the formulation of the campaign not only informed by the historic donor behavior but also the prospects.

E. Implications for Non-Profit Practice
Practically, effective execution of clustering demands investments in data infrastructure, analytical resources, and employee education [10]. These requirements can be overwhelming for smaller organizations because of resource constraints. Nevertheless, with increased access to open-source software and cloud-based solutions, financial limitations are lowered, and sophisticated analytics become more accessible.

VI. CONCLUSION
This report has discussed the use of clustering algorithms to optimize donor segmentation as a strategy to enhance donor retention in the non-profit sector. The analysis has shown that clustering facilitates the identification of relevant donor types that, consequently, direct the creation of more accurate and precise fundraising strategies. Through k-means, hierarchical clustering, and DBSCAN, it became possible to showcase the variety of donor behaviors and emphasize the need to shape engagement strategies to various groups like loyal supporters, high-value champions, occasional donors, and at-risk individuals. Data quality and completeness are vital for generating reliable insights, while algorithm selection influences the stability and interpretation of results. Ethical concerns, including privacy and potential bias, must also be addressed to sustain donor trust and ensure fairness in engagement. Future research could explore integrating clustering with predictive modelling and real-time analytics to create adaptive systems that respond dynamically to donor behavior.

REFERENCES:
[1] L. Heumos *et al.*, “Best practices for single-cell analysis across modalities,” *Nature Reviews Genetics*, vol. 24, no. 8, pp. 1–23, Mar. 2023, doi: https://doi.org/10.1038/s41576-023-00586-w.
[2] D. Lähnemann *et al.*, “Eleven grand challenges in single-cell data science,” *Genome Biology*, vol. 21, no. 1, Feb. 2020, doi: https://doi.org/10.1186/s13059-020-1926-6.

[3] A. De Bruyn, V. Viswanathan, Y. S. Beh, J. K.-U. Brock, and F. Von Wangenheim, “Artificial Intelligence and marketing: Pitfalls and opportunities,” *Journal of Interactive Marketing*, vol. 51, no. 1, pp. 91–105, Jun. 2020.
[4] T. L. Nguyen, S. Pradeep, R. L. Judson-Torres, J. Reed, M. A. Teitell, and T. A. Zangle, “Quantitative Phase Imaging: Recent Advances and Expanding Potential in Biomedicine,” *ACS Nano*, vol. 16, no. 8, pp. 11516–11544, Aug. 2022, doi: https://doi.org/10.1021/acsnano.1c11507.
[5] H. Lambers, “Phosphorus Acquisition and Utilization in Plants,” *Annual Review of Plant Biology*, vol. 73, no. 1, Dec. 2021, doi: https://doi.org/10.1146/annurev-arplant-102720-125738.
[6] Nataliia Letunovska and V. Kazymirova, “USING MARKETING TECHNIQUES TO PROMOTE HEALTHY LIFESTYLES AMONG VARIOUS SEGMENTS OF PEOPLE,” *Ìnfrastruktura rinku*, no. 72, Jan. 2023, doi: https://doi.org/10.32782/infrastruct72-8.
[7] C. B. Jackson, M. Farzan, B. Chen, and H. Choe, “Mechanisms of SARS-CoV-2 Entry into Cells,” *Nature Reviews Molecular Cell Biology*, vol. 23, no. 1, pp. 1–18, Oct. 2021, doi: https://doi.org/10.1038/s41580-021-00418-x.
[8] X. Li *et al.*, “Diselenide as a Dual Functional Mechanophore Capable of Stress Self-Reporting and Self-Strengthening in Polyurethane Elastomers,” *CCS Chemistry*, vol. 5, no. 4, pp. 925–933, Apr. 2022, doi: https://doi.org/10.31635/ccschem.022.202201874.
[9] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, “Transformers in Vision: A Survey,” *ACM Computing Surveys*, Jan. 2022, doi: https://doi.org/10.1145/3505244.
[10] Z. Z. Lim, S. Tay, F. Wang, S. Ling, Y. L. Lim, and Y. Wang, “EE245 Indirect Cost of Spinal Muscular Atrophy (SMA) in Singapore,” *Value in Health*, vol. 26, no. 12, p. S98, Dec. 2023, doi: https://doi.org/10.1016/j.jval.2023.09.512.
[11] S. Y. Chong *et al.*, “Injectable liposomal DHA alleviates atherosclerosis progression and enhances plaque stability,” *Atherosclerosis*, vol. 379, p. S7, Aug. 2023, doi: https://doi.org/10.1016/j.atherosclerosis.2023.06.702.
[12] C.-W. Hsu, Y.-L. Chang, T.-S. Chen, T.-Y. Chang, and Y.-D. Lin, “Who Donates on Line? Segmentation Analysis and Marketing Strategies Based on Machine Learning for Online Charitable Donations in Taiwan,” *IEEE Access*, vol. 9, pp. 52728–52740, 2021, doi: https://doi.org/10.1109/access.2021.3066713.
[13] T. M. Shami, A. A. El-Saleh, M. Alswaitti, Q. Al-Tashi, M. A. Summakieh, and S. Mirjalili, “Particle Swarm Optimization: A Comprehensive Survey,” *IEEE Access*, vol. 10, pp. 10031–10061, 2022, doi: https://doi.org/10.1109/ACCESS.2022.3142859.
[14] X.-X. Zeng et al., “An Amphipathic Ionic Sieve Membrane for Durable and Dendrite-Free Zinc-Ion Batteries,” Renewables, vol. 2, no. 1, pp. 52–60, Feb. 2024, doi: https://doi.org/10.31635/renewables.024.202300045.
[15] S.-J. Bao, H.-N. Zhang, and G.-X. Jin, “Stereoselective Self-Assembly of Chiral Metalla[2]Catenanes and Lemniscular Macrocycles,” CCS Chemistry, pp. 1–11, Jan. 2024, doi: https://doi.org/10.31635/ccschem.024.202303525.