4/29/2024

Daisy Ipatzi Bello

University of essex online

Implementing Machine Learning tools and/or techniques in Customer Profiling

*A literature Review*

**Abstract**

This literature review evaluates the use of K-means clustering for customer profiling in B2B ecommerce. Analysing a range of studies from academic databases, the review explores the algorithm’s potential and challenges in the B2B context. Findings indicate that while K-means offers valuable insights for customer segmentation, its application is limited by the complexity of B2B datasets. The review highlights the need for further research to optimize K-means for B2B ecommerce, providing both theoretical and practical contributions to the field.

**Introduction**

Customer profiling assumes a central role within Business-to-Business (B2B) ecommerce, intricately shaping targeted marketing strategies for precision and efficacy. Unlike Business-to-Consumer (B2C) contexts, B2B transactions involve a thinner yet diverse customers, demanding a nuanced understanding of customer behaviours and preferences. Customer profiling serves as the requirement, empowering businesses to tailor their marketing endeavours to the distinctive needs of various client segments, thereby enhancing customer engagement and driving sales.

Peppers and Rogers (2010) highlight the significance of adopting a customer-centric strategy, emphasising its potential to foster loyalty, facilitate cross-selling, and ultimately bolster revenue. Conversely, a failure to adequately address customer needs may precipitate defection to competitors, underscoring the importance of retention strategies. This is particularly pertinent given the substantial expense associated with acquiring new customers compared to serving existing ones (Kotler et al., 2013).

At the core of customer segmentation in B2B ecommerce lies the utilisation of data-driven methodologies, with K-means clustering emerging as a prominent tool. This algorithm, rooted in machine learning and data analytics, offers promise in segmenting B2B customers based on a synthesis of purchasing behaviours and demographic characteristics.

This literature review aims to critically gain understanding of K-means clustering in profiling B2B ecommerce customers, as framed by the research question:

*"How effective is K-means clustering in segmenting B2B ecommerce customers based on purchasing behaviours and demographic characteristics for improving targeted marketing strategies?"*

This study adheres to the IMRAD structure, comprising Introduction, Methods, Results, and Discussion sections (Aarhus University, 2022). The Research Strategy section delineates the methodology utilised for conducting the literature review. Following this, the Review of Literature section scrutinises the current state of the field, specifically focusing on the research question at hand. Lastly, the Discussion and Conclusion section integrates the findings, analyses their implications, and offers insights into the necessity of addressing the research question.

**Research Strategy**

For the research strategy, this study will adopt a structured approach to gather relevant scholarly materials on B2B ecommerce customer segmentation and K-means clustering. Leveraging my experience in marketing and familiarity with the topic, literature selection commenced over the past month. The search focused on renowned academic databases and search engines, such as ACM Digital Library, IEEE Xplore Digital Library, the Essex Online Library, ScienceDirect and Google Scholar. Search terms will be tailored to encompass key concepts, and filters will be applied to refine search results, such as credibility.

**`**

**Review of Literature**

This review critically examines a collection of studies exploring the efficacy of K-means clustering, a widely adopted algorithm, in profiling customers. Through an analysis of these studies, this review aims to extract insights into the applicability and effectiveness of K-means clustering in the context of B2B customer segmentation, thereby offering valuable guidance to businesses navigating the complexities of targeted marketing in the B2B landscape.

K-Means is an iterative algorithm that divides unlabelled data into K characteristic clusters. The algorithm alternates among 2 steps: assigning every information factor to the closest cluster and recalculating the midpoint of every cluster. The algorithm is done once the assignment of situations to clusters no longer changes. The midpoint of a cluster is described as the suggest of the information points, a prototype of the cluster. Information points are assigned to the cluster that has the nearest suggest (Moshkovitz et al.2020).

Azevedo and Gartner (2020) provide illuminating insights into market concentration and competition within the domestic credit market. While their study focuses on a different industry, the principles of market segmentation and competition dynamics they discuss offer valuable parallels for understanding customer segmentation in B2B ecommerce. By examining market dynamics in diverse sectors, businesses can glean insights into effective segmentation strategies and competitive positioning.

Borlea et al. (2021) introduce a unified form of fuzzy C-means and K-means algorithms, presenting potential refinements to clustering techniques for customer segmentation. Their study contributes to the ongoing discourse on algorithmic advancements in clustering methods, shedding light on avenues for improving the accuracy and robustness of customer segmentation models. While the application of these techniques to B2B ecommerce requires further investigation, the theoretical advancements proposed by Borlea et al. offer promising directions for future research.

Essayem et al. (2022) directly address the research question by exploring customer clustering based on RFM (Recency, Frequency, Monetary) features using the K-means algorithm. Through their application of K-means clustering to segment B2B ecommerce customers, Essayem et al. offer valuable insights into the practical implications of this approach. By leveraging RFM data, businesses can identify distinct customer segments with similar purchasing patterns, enabling targeted marketing efforts tailored to specific customer needs and preferences.

Govender and Sivakumar (2020) offer a comprehensive review of the application of k-means and hierarchical clustering techniques for the analysis of air pollution. While the focus of their study diverges from B2B ecommerce, the methodological insights they provide hold relevance for customer segmentation studies. By adapting clustering techniques developed in other domains, businesses can explore novel approaches to segmenting their customer base, potentially enhancing the accuracy and granularity of segmentation models.

Huang et al. (2021) propose a robust deep k-means method for data clustering, aiming to improve clustering accuracy in B2B ecommerce settings. While their study primarily focuses on technical advancements in clustering algorithms, the implications for customer segmentation in B2B ecommerce are noteworthy. By exploring innovative clustering techniques, businesses can potentially uncover hidden patterns and relationships within their customer data, facilitating more precise segmentation and targeted marketing strategies.

Li et al. (2021) employ K-means clustering and adaptive particle swarm optimization algorithm for customer segmentation, introducing a novel approach to segmenting B2B ecommerce customers based on purchasing behaviours. Through their empirical findings, Li et al. offer insights into the effectiveness of hybrid clustering techniques for customer segmentation. By leveraging adaptive optimization algorithms, businesses can potentially enhance the accuracy and efficiency of their segmentation models, leading to more effective targeted marketing campaigns.

Mensouri et al. (2022) focus on K-means customers clustering based on RFMT (Recency, Frequency, Monetary, Tenure) and score satisfaction analysis, providing insights into the integration of additional variables for refined customer segmentation in B2B ecommerce. Through their exploration of multi-dimensional segmentation approaches, Mensouri et al. underscore the importance of considering diverse factors in customer segmentation. By incorporating tenure and satisfaction scores into the segmentation process, businesses can gain a holistic understanding of their customer base, enabling tailored marketing strategies that resonate with customer preferences.

Mussabayev et al. (2022) explore the application of K-means for big data clustering, offering insights into scalability and efficiency considerations for clustering algorithms in large-scale B2B ecommerce datasets. While their study primarily focuses on technical aspects of clustering algorithms, the implications for B2B customer segmentation are noteworthy. Scalability is a critical factor for businesses dealing with large and dynamic datasets, and understanding the scalability of clustering algorithms can inform strategic decisions around data management and segmentation.

Nedyalkova et al. (2021) propose combinatorial K-means clustering as a machine learning tool for diabetes mellitus type 2, showcasing the versatility of K-means clustering in diverse domains and its potential applicability in B2B ecommerce customer segmentation. Through their exploration of K-means clustering in a healthcare context, Nedyalkova et al. highlight the adaptability of clustering techniques to different data types and domains. By drawing parallels between healthcare and B2B ecommerce, businesses can glean insights into innovative approaches to customer segmentation.

Nouraei et al. (2022) compare unsupervised machine learning approaches for cluster analysis in heart failure with preserved ejection fraction, offering insights into the comparative effectiveness of clustering algorithms, including K-means, in segmenting heterogeneous datasets. While their study focuses on a different domain, the comparative analysis of clustering algorithms offers valuable insights for B2B ecommerce customer segmentation. By identification the strengths and weaknesses of various clustering techniques, businesses can make informed decisions about the most suitable approach for segmenting their customer base.

Srivastava et al. (2023) present a case study of advanced data analytics in the telecommunications industry, highlighting the practical applications of data analytics techniques, including clustering algorithms, in real-world business contexts. While their study primarily focuses on telecommunications, the insights into data analytics methodologies hold relevance for B2B ecommerce. By understanding how clustering algorithms are applied in diverse industries, businesses can gain inspiration for innovative approaches to customer segmentation in B2B ecommerce.

Vohra et al. (2020) utilize self-organizing maps and K-means clustering based on the RFM model for customer segmentation in the online retail business, offering insights into hybrid clustering approaches for customer segmentation in ecommerce settings. Through their exploration of hybrid clustering techniques, Vohra et al. demonstrate the potential for integrating multiple clustering algorithms to enhance customer segmentation accuracy. By combining the strengths of different clustering methods, businesses can develop more robust segmentation models that capture the nuances of their customer base.

Zare and Emadi (2020) focus on determining customer satisfaction using an improved K-means algorithm, providing insights into the role of clustering techniques in understanding customer preferences and satisfaction levels in B2B ecommerce. Through their examination of customer satisfaction data, Zare and Emadi highlight the value of clustering techniques in uncovering actionable insights for businesses. By applying clustering algorithms to customer satisfaction data, businesses can identify patterns and trends that inform targeted marketing strategies and drive customer satisfaction.

**Identifying Gaps in Research**

Despite the wealth of studies examining the application of clustering algorithms in various domains, there are notable gaps in the literature concerning the specific application of K-means clustering in Business-to-Business (B2B) ecommerce customer segmentation. While existing research provides valuable insights into algorithmic advancements, methodological approaches, and practical applications of clustering techniques, there is a scarcity of studies that directly address the intricacies of B2B customer segmentation.

Furthermore, the available literature often presents contrasting views on the effectiveness of K-means clustering, with some studies highlighting its potential for accurate segmentation and targeted marketing, while others raise concerns about its limitations in handling complex B2B datasets and capturing the nuanced relationships between businesses.

Moreover, the challenge of finding relevant research in the B2B context is exacerbated by the multifaceted nature of B2B transactions, which involve multiple individuals within an organization contributing to the purchasing decision-making process. As such, there is a pressing need for more focused research that delves into the unique challenges and opportunities of B2B customer segmentation, offering actionable insights for businesses operating in this complex ecosystem.

Additionally, the limited number of documented use cases of K-means clustering specifically tailored for B2B ecommerce underscores the need for more comprehensive studies exploring feature selection and real-time implementation of this model. While K-means clustering has shown promise in various domains, its application in B2B ecommerce remains relatively underexplored. Understanding the selection of features relevant to B2B transactions and the practical challenges associated with implementing K-means clustering in real-time ecommerce environments is crucial for bridging this gap between theory and practice.

**Discussion and Conclusion**

In this literature review, we have explored the effectiveness of K-means clustering in segmenting Business-to-Business (B2B) ecommerce customers. Through an analysis of a diverse range of studies, it becomes evident that while K-means clustering holds promise as a tool for customer segmentation, its application in the B2B context remains not frequently used. While some studies demonstrate the potential of K-means clustering to accurately segment B2B customers and inform targeted marketing strategies, others highlight its limitations in handling complex datasets and capturing the nuanced relationships between businesses.

Future studies should focus on addressing key research gaps, such as feature selection, model evaluation, and real-time implementation of clustering algorithms in B2B ecommerce environments. By delving into these areas, researchers can contribute to advancing the effectiveness and applicability of K-means clustering or some other customer profiling in B2B customer segmentation, thereby empowering businesses to make data-driven decisions and optimize their marketing strategies.

As someone deeply involved in a B2B company, I have firsthand experience of the challenges associated with implementing customer profiling in a complex business model where buyers are not necessarily the end customers. Therefore, I am particularly interested in delving deeper into this topic to explore how clustering algorithms like K-means can be tailored to address the unique challenges of B2B ecommerce. By gaining a better understanding of feature selection, model evaluation, and real-time implementation strategies, I aim to contribute to the development of more effective customer segmentation strategies that drive business growth and enhance customer satisfaction in the B2B landscape.

**References**

Aarhus University (2022). *IMRAD structure*. [online] studypedia.au.dk. Available at: https://studypedia.au.dk/en/formal-requirements/imrad-structure [Accessed 27 Apr. 2024].

Azevedo, M. de A. and Gartner, I.R. (2020). Concentração e Competição no Mercado de Crédito Doméstico. *Revista de Administração Contemporânea*, [online] 24, pp.380–399. doi:https://doi.org/10.1590/1982-7849rac2020190347.

Borlea, I.-D., Precup, R.-E., Borlea, A.-B. and Iercan, D. (2021). A Unified Form of Fuzzy C-Means and K-Means algorithms and its Partitional Implementation. *Knowledge-Based Systems*, 214, p.106731. doi:https://doi.org/10.1016/j.knosys.2020.106731.

Essayem, W., Bachtiar, F.A. and Priharsari, D. (2022). *Customer Clustering Based on RFM Features Using K-Means Algorithm*. [online] IEEE Xplore. doi:https://doi.org/10.1109/CyberneticsCom55287.2022.9865572.

Golash-Boza, T. (2022). *How to Write a Literature Review: Six Steps to Get You from Start to Finish*. [online] www.wiley.com. Available at: https://www.wiley.com/en-us/network/publishing/research-publishing/writing-and-conducting-research/writing-a-literature-review-six-steps-to-get-you-from-start-to-finish [Accessed 2 Apr. 2024].

Govender, P. and Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, 11(1), pp.40–56. doi:https://doi.org/10.1016/j.apr.2019.09.009.

Huang, S., Kang, Z., Xu, Z. and Liu, Q. (2021). Robust deep k-means: An effective and simple method for data clustering. *Pattern Recognition*, 117, p.107996. doi:https://doi.org/10.1016/j.patcog.2021.107996.

Kotler, P. and Armstrong, G. (2014). *Principles of marketing*. Upper Saddle River, N.J.: Pearson.

Li, Y., Chu, X., Tian, D., Feng, J. and Mu, W. (2021). Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm. *Applied Soft Computing*, 113, p.107924. doi:https://doi.org/10.1016/j.asoc.2021.107924.

Lloyd, C. (2017). *LITERATURE REVIEWS FOR SOCIOLOGY SENIOR THESES*. [online] Available at: https://socthesis.fas.harvard.edu/files/socseniorthesis/files/pres-litreview.pdf [Accessed 2 Apr. 2024].

Mensouri, D., Azmani, A. and Azmani, M. (2022). K-Means Customers Clustering by their RFMT and Score Satisfaction Analysis. *International Journal of Advanced Computer Science and Applications*, 13(6). doi:https://doi.org/10.14569/ijacsa.2022.0130658.

Moshkovitz, M., Dasgupta, S., Rashtchian, C. and Frost, N., 2020, November. Explainable k-means and k-medians clustering. In International conference on machine learning (pp. 7055-7065). PMLR. mlr.press

Mussabayev, R., Mladenovic, N., Jarboui, B. and Mussabayev, R. (2022). How to Use K-means for Big Data Clustering? *Pattern Recognition*, p.109269. doi:https://doi.org/10.1016/j.patcog.2022.109269.

Nedyalkova, M., Madurga, S. and Simeonov, V. (2021). Combinatorial K-Means Clustering as a Machine Learning Tool Applied to Diabetes Mellitus Type 2. *International Journal of Environmental Research and Public Health*, 18(4), p.1919. doi:https://doi.org/10.3390/ijerph18041919.

Nouraei, H., Nouraei, H. and Rabkin, S.W. (2022). Comparison of Unsupervised Machine Learning Approaches for Cluster Analysis to Define Subgroups of Heart Failure with Preserved Ejection Fraction with Different Outcomes. *Bioengineering*, 9(4), p.175. doi:https://doi.org/10.3390/bioengineering9040175.

Peppers, D. and Rogers, M. (2010). *Managing Customer Relationships*. John Wiley & Sons.

Practical Data Science (2021). *A quick guide to customer segmentation for B2B e-commerce*. [online] practicaldatascience.co.uk. Available at: https://practicaldatascience.co.uk/data-science/a-quick-guide-to-customer-segmentation-for-b2b-e-commerce#google\_vignette [Accessed 27 Apr. 2024].

Sembiring Brahmana, R.W., Mohammed, F.A. and Chairuang, K. (2020). Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 11(1), p.32. doi:https://doi.org/10.24843/lkjiti.2020.v11.i01.p04.

Srivastava, S., Giri Gundu Hallur and Mukitm, A. (2023). Advanced-data analytics in telecommunications industry: A case study of accenture. *Nucleation and Atmospheric Aerosols*. [online] doi:https://doi.org/10.1063/5.0170732.

Vohra, R., Pahareeya, J., Hussain, A., Ghali, F. and Lui, A. (2020). Using Self Organizing Maps and K Means Clustering Based on RFM Model for Customer Segmentation in the Online Retail Business. *Intelligent Computing Methodologies*, pp.484–497. doi:https://doi.org/10.1007/978-3-030-60796-8\_42.

Zare, H. and Emadi, S. (2020). Determination of Customer Satisfaction using Improved K-means algorithm. *Soft Computing*, 24(22), pp.16947–16965. doi:https://doi.org/10.1007/s00500-020-04988-4.