Manuscript Submitted	15.5.2024
Accepted	11.6.2024
Published	30.6.2024

Exploring Hypermarket Product Segmentation using Machine Learning Approach in Muscat City, Oman

Marwan Abdullah Hasan Al Kubati, Nurazean Maarop, Norshaliza Kamaruddin, Ganthan Narayana Samy, Norziha Megat Mohd Zainuddin, Roslina Mohammad

Faculty of Artificial Intelligence
Universiti Teknologi Malaysia
hasan-20@graduate.utm.my, nurazean.kl@utm.my, norshaliza.k@utm.my, ganthan.kl@utm.my, norziha.kl@utm.my, mroslina@utm.my

DOI: https://doi.org/10.53840/myjict9-1-187

Abstract

The Omani retail industry faces various issues, including identifying patterns in customer behavior and categorizing sold products, analyzing the value of products, and segmenting these products for marketing purposes. In addition, the wide range of products and the differences in consumer purchasing patterns across various product categories, as well as the challenges in precisely monitoring and assigning sales to customers. This research aims to explore the product segmentation at one hypermarket located in Muscat, Sultanate of Oman. Based on consumers' purchase histories, recency frequency monetary value (RFM) Analysis was used to project a product sales trend and to provide recommendations to stakeholders for how to enhance digital platforms. Using methods from unsupervised machine learning, this study constructed models employing K-means and Fuzzy C-means algorithms for hard clustering, and the density-based clustering algorithm (DBSCAN) for soft clustering. Product segmentation procedures were performed to gauge the three algorithms' relative performance. The findings revealed that the DBSCAN method had the best performance, scoring 0.89 across all three clusters, while the K-means algorithm scored 0.724 and the Fuzzy C-means strategy scored 0.702. The results may be used to shed light on client buying habits and provide data-driven suggestions for enhancing stakeholders' digital platforms. In preventing issues like inventory loss, stockouts, and overstock, businesses might benefit from delivering customized suggestions based on consumers' purchasing behavior history. Improved marketing, stock control, and customer satisfaction are all possible benefits from this study.

***Keywords*:** *Product Segmentation, RFM Analysis, K-means, Machine Learning, Clustering*

1. Introduction

Hypermarkets, the widely favored self-service establishments, offer a diverse range of items and foods conveniently located in a single location. In 1961, the Great Britain firm constructed the first Hypermarket in Bruges, consisting of three Supermar-kets located in Auderghem and Anderlecht. The Hypermarket was branded as Super Bazar and had a local area of 9100 m² (Grimmeau, 2013). The digital revolution, which began in the 1950s, introduced various groundbreaking inventions such as the Internet, social media, mobile phones, apps, cloud computing, big data, and e-commerce, which refers to the electronic transactions of goods and services (Alghanam et al., 2022). These advancements, along with the consumerization of IT, have led to a comprehensive transformation in products, services (Vehmas et al., 2014) marketing, and global business processes. The retail industry is likewise impacted

by these impacts, which have produced notable advancements in marketing (Hengsberger, 2018). Machine learning (ML), a component of Artificial intelligence (AI) (Ma and Son, 2020) is an essential tool for constructing prediction and prescription models, as well as analyzing consumer behavior (CB) patterns through clustering. It offers a comprehensive method-ology for extracting valuable information from data. Furthermore, it provides valuable business intelligence, automates tasks, and enhances the capabilities of their system. In addition, the majority of academics contend that automation and monitoring have a significant impact on all facets of the business process (Ma and Son, 2020). While machine learning aids stakeholders in comprehending their customers' needs and identifying their behavioral patterns, it is ultimately the stake-holders' responsibility to choose how to address consumers based on these findings and how to formulate their business strategies accordingly. Over the past decade, recommendation systems (RSS) have emerged as a very effective use of machine learning (ML). RSS was created inside the Digital Platform (DP) to have an impact on consumers' behavior, namely in terms of their economic choices and decision-making processes (Nguyen, 2021)

Assisting customers in making purchasing decisions at hypermarket DPs can enhance the satisfaction of existing clients and perhaps attract new customers (Lee and Hosanagar, 2019). Nevertheless, this procedure is not straightforward. The recommended list (RL) may have a contrary effect if the items on it do not align with the customer's interests. This ML strategy is not arbitrary; rather, it relies on complex algorithms and machine learning models (Ma and Son, 2020). There are several factors that govern these activities that must occur prior to selecting that RL. These metrics may be derived from the hypermarket data through the utilization of data mining technologies. Anticipating the consumer's preference is seen as an aspiration for organizations, aiding in comprehension. Con-structing a machine learning model necessitates significant focus and the identification of appropriate parameters. However, the impact of these models on customers varies from one client to another, and even from one city to another (Medrano et al., 2015). The Omani retail business has several issues, including identifying patterns in customer behavior and categorizing sold items, evaluating the value of products, and segmenting them for marketing strategies. Furthermore, the wide array of products and the differences in how customers purchase them in various product categories have led to challenges in precisely monitoring and assigning sales to specific customers. Therefore, the objective of this research is to conduct segmentation analysis by constructing machine learning models using appropriate algorithms such as K-means, FCM, and DBSCAN.

2. Literature Review

2.1 E- Commerce and Customer Behavior in Hypermarket

Customer behavior might vary across different countries and even within cities (Nguyen, 2021). Researchers have detected impulsive purchase behavior (IPB) for the first time, which can directly impact buying in hypermarkets both in-store and online. According to Chauhan et al. (2023), IPB influences a significant proportion of buying choices, ranging from 40% to 80%. IPB, or impulse buying, refers to the act of purchasing products without any deliberation. Businesses utilize visual merchandising techniques, including enticing displays, shop layouts, pricing strategies, sales promotions, packaging, and diverse product categories, to provide unanticipated stimulation (Amos et al., 2014). Furthermore, IPB is influenced by two distinct categories of influences, namely external and internal. clients in Muscat city, including employees and workers facing time constraints, exhibit a preference for e-commerce (Erjavec and Manfreda, 2022). Additionally, vulnerable clients with health issues may have a propensity towards online purchasing (Drenik, 2021). In addition, non-native workers of other ethnic backgrounds, who are unfamiliar with the surroundings, often choose for internet sales rather than in-person buying.

Recommender lists predict users' preferences for goods and suggest items that users are likely to purchase. Recommendation methods are often categorized into three distinct groups: collaborative filtering, content-based filtering, and hybrid models (Hu et al., 2019). The analysis in a hypermarket

relies on the behavior of customers, including their choices and behaviors that are similar to those of other customers. These parameters are crucial for the development of digital platforms used for onsite shopping in hypermarkets (Dobre and Milovan-Ciuta, 2015).

2.2 Machine Learning and Data Mining Techniques

Clustering analysis is a traditional descriptive method used for market segmentation. Unsupervised machine learning analysis can be used to implement this technique, which is commonly employed to understand the characteristics of buyers and markets, as well as to discover interesting patterns in the business domain. This approach can also be used to analyze time stamps (Bang et al., 2015) and identify the most profitable buyers (Ramon-Jeronimo, 2009). This approach relies on dividing the data into a certain number of segments, in order to identify the patterns related to the qualities needed for segment analysis. Various methods employ clustering methodologies, including K-means, Affinity Propagation, DBSCAN, Mini-Batch, Mean Shift, Birch, optics, and spectral clustering (Brownlee, 2020). Although the goal is the same, the algorithm employs distinct procedures. Provide a more precise segmentation of online purchases by analyzing consumers' behaviors using real sales data. The K-means algorithm, a method for grouping data, is easily interpretable for dataset analysis. Employing several types of data may be useful, however determining the appropriate number of clusters can be challenging. Nevertheless, K-means clustering remains necessary in big data applications because to the high cost of transmission overhead (Zhang et al., 2022)

2.3 Hard Clustering by Using the K-Means s Algorithm

The K-means algorithm, proposed by Alsabti et al. (1997), is a hard clustering approach used to construct machine learning models. It is designed for unsupervised data and aims to create η partitions for the data observations. The algorithm selects a single point from each cluster and designates it as the centroid, which represents the center of the cluster. The subsequent procedure involves calculating the distance between each observation and the centroids, as well as determining the distance to the nearest centroid in order to assign it to the corresponding cluster (Fahim et al., 2006). Quantitative data clustering, particularly sales data, can be advantageous in identifying patterns in customer behavior and sales data within the retail sector. Prior research has consistently recognized the RFM model as the optimal approach for improving consumer understanding and clustering (Mahfuza et al., 2022). In their study, Gustriansyah et al. (2020) utilized K-means algorithms to optimize clustering in RFM analysis in the Indonesian market. They found that the optimum silhouette score, which measures the quality of the clustering, was 0.49 when the number of clusters (k) was set at 4. This result was obtained by considering eight different performance metrics. Brahmana et al. (2020) developed a customer segmentation using the RFM model and employed the K-Means and K-Medoids algorithms on a dataset consisting of 334,641 transactions. They determined that the optimal performance value for the Silhouette score was 0.330.

2.4 Soft Clustering by Fuzzy C-Mean Algorithm

Several scholars believe that K-means has difficulties when dealing with data that exhibits variation in terms of density and sizes. Nevertheless, the utilization of K-means remains widely favored. The FCM algorithm has been enhanced to effectively accomplish the objective of machine learning, particularly in clustering unlabeled data (Bezdek, 2013). The FCM algorithm is widely regarded as one of the soft clustering approaches. On the other hand, soft approaches encompass techniques such as fuzzy c-mean, which calculates the chance of data being assigned to a fuzzy cluster. This means that a single data point might potentially belong to many clusters Idowu et al. (2019). Furthermore, FCM assigns membership to each node in the cluster. However, the K-means algorithm is used to assign data points to distinct clusters. In addition, a study utilized the fuzzy c-mean algorithm along with association rules to create the initial cluster, which served as the machine learning model. The researchers argued that traditional clustering methods are inadequate for defining clusters as modules in product design (Moon et al.,2006)

2.5 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

DBSCAN is a non-parametric density-based technique that was created in 1996 by Ester et al. (1999) The DBSCAN algorithm operates by clustering data points that are in close proximity to one other. DBSCAN method distinguishes itself from K-means and fuzzy C-means algorithms by not necessitating a predefined number of clusters. However, it just needs one input parameter that represents the degree of fuzziness. Brahmana et al. (2020) used the DBSCAN algorithm to a retail business in Indonesia. They conducted customer segmentation analysis using eight performance variables and found that the DBSCAN silhouette score was 0.910, indicating high performance. Furthermore, Shirole et al. (2021) conducted a study on client segmentation in the Indian market, employing the k-means algorithm and DBSCAN. Although the data spanned a period of thirteen months, the DBSCAN algorithm achieved a high performance of 0.442, as reported by Shirole et al. (2021). Moreover, the study conducted by Alsahafi et al (2022) examined the use of unsupervised learning and data clustering in Saudi online stores to analyze client behavior using the same techniques. Nevertheless, the silhouette value for K-means was 0.1699, while the DBSCAN silhouette score was 0.206, indicating a lower score.

2.6 Parameters Used in Previous Machine Learning Models

The RFM model was developed in 1994 to help the organization to know better for their data and score their business. then it has become more popular to segment the clients and products. There is method to calculate the RFM score depends on three factors that control RFM models factors using different algorithms to cluster that data. RFM models used in many previous researchers to clusters of customers in different categories depend on their behavior as shown in Table 1 showing some research with algorithms and attributes they had used in their research. several studies about customer or product clustering. Indifferent fields or research sectors such as products, CB clustering, and ML by the implementation of K-means s, Fuzzy c-mean, and DBSCAN algorithms. Previous study on customer behavior (CB) analytics in Muscat city, Sultanate of Oman, has several limitations. These constraints arise from the fact that each city has its own unique CB and many characteristics that influence it. Consequently, the segmentation of items into distinct groups is dependent on the specific CB of each city. Currently, there are only two analytical studies conducted on consumer behaviors in CBs inside Muscat hypermarkets. However, not enough study has been done on creating a machine learning model that can correctly predict how customers be-have.

Table 1: Parameters used in previous ML models

Author	Title	Algorithm	Parameter
(Tahiri et al., 2019)	An intelligent shopping list based on the application of partitioning and machine learning algorithms	K-means, C-means	User-id, Products_id, Categories, Sold unit
(Zuo et al., 2016)	Prediction of Consumer Purchasing in a Grocery Store Using Machine Learning Techniques	SVM, K-means.	Customers' Age, Gender, Income, Product Price, Sale Promotion
(Bruestle et al., 2019)	Prediction of consumer purchasing in a grocery store using machine learning techniques	Hierarchical, K-means s clustering	Product_id, categories, quantities. amount.
(Kseniia, 2018)	Market basket visualization for hypermarket with the use of big data analytics	K-means clustering algorithm.	Transactions_ID, Product, Quantities, Amount, Date

(Alghanam et al., 2022)	Data Mining Model for Predicting Customer Purchase Behaviour in E-Commerce Context.	K-means s, C4.5, J48, CS-MC4, MLR	Customer_id, Gender, Order-id, Categories _name, Unit price, Quantity Discount
(Koofan and Kaleem, 2019)	Analyze and Enhance Sales in Lulu Supermarket using Data Mining Technology	Apriori algorithms. cluster-based association rule (CBAR), matrix-based scheme (MBS), Dashbased scheme (HBS).	Order_no, Product, Price, Customer_id

3. Methodology

This study applied K-means, FCM, and DBSCAN algorithms along with Recency Frequency Monetary analysis approach. Several steps were taken in the data analysis approach. Firstly, exploratory data analysis (EDA) process was conducted on the data sets to explore and treat the missing values, as well as to determine the attributes which were useful to achieve the analysis goal. Secondly, we designed the clustering process and developed the Machine Learning RFM Model using the two types of algorithms; centroid-based clustering algorithms (K-means and FCM) and density-based algorithm (DBSCAN). Lastly, we evaluated the Machine Learning models by finding which algorithm has the higher performance.

The dataset which has been supplied consists of 1,618,599 transactions represent the sales for one of the hypermarkets in Muscat, Sultanate of Oman during six months. The datasets files include varying numbers of characteristics owing to the different files obtained from the source. Some attributes were not used in either designing the dashboard or for clustering using Python. Only the relevant attributes were used. Table 2 shows the attributes and types of datasets.

Table 2: Hypermarket dataset attributes and types.

Dataset		
Attribute Name	**ATT.Detail**	**Data Type**
StoreNo	Store No.	String
POSTerminalNo	POS Terminal No.	String
TransactionNo	Transaction No.	Numeric
LineNo	Line No.	Numeric
ReceiptNo	Receipt No.	Numeric
ItemNo	Item No.	Numeric
ItemCategoryCode	Item Category Code	String
ProductGroupCode	Product Group Code	String
Price	Omani R	Numeric
Quantity	Pcs	Numeric
Date	date	Date
Time	time	Date
TransactionCode	Transaction Code	String
ItemNumberScanned	Item Number Scanned	String
KeyboardItemEntry	Keyboard Item Entry	String
PriceinBarcode	Price in Barcode	String
PriceChange	Promotion	String
WeightManuallyEntered	Weight Manually Entered	String
LinewasDiscounted	Line was Discounted	String
ScaleItem	Scale Item	String
WeightItem	Weight Item	String
ReturnNoSale	Return No Sale	String
ItemCorrectedLine	Item Corrected Line	String
TypeofSale	Type of Sale	String
LinkedNo.notOrig	Linked No. not Orig.	String
Orig.ofaLinkedItemList	Orig. of a Linked Item List	String
StaffID	Staff ID	Numeric
ItemPostingGroup	Item Posting Group	String
UnitofMeasure	Unit of Measure	String
CouponDiscount	Coupon Discount	Numeric
CouponAmt.ForPrinting	Coupon Amt. For Printing	Numeric
ReplicationCounter	Replication Counter	Numeric
Description		String
Description2		String

The given data from the hypermarket sales database contains the previous months sales, with different attributes which tagged to demographic information and preferences and with a mixed type of data such as string, integer, float, boolean, times, and date alongside the important data such as products, categories, detail for the product, price and quantity The given dataset were analyzed to find the pattern of consumers' purchase behaviors in order to improve data processing and point of sales.

With regard to clustering, we employed K-means, FCM, and DBSCAN algorithms and let them run on Jupyter Notebook and Google Colab tool. Earlier, the result on the literature review suggested the potential attributes which may be used to build machine learning model and which have direct influence on the sales. The figure 1 illustrates the major steps to conduct this part.

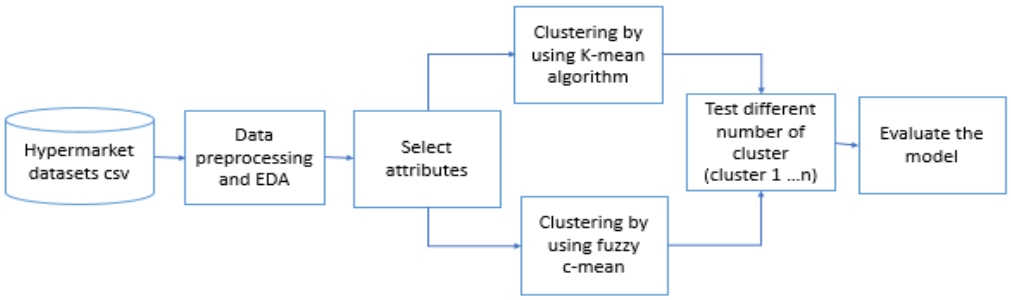

Figure 1: Clustering Procedure

In addition, this study utilized the DBSCAN method to compare the findings without considering the outlier data. DBSCAN is a robust technique specifically designed for handling outlier data, and requires two parameters: eps (epsilon) and minPts. The eps represent the maximum radius of the neighborhood from a data point, and min sample is the minimum number of data points required to form a dense region (or a cluster). Firstly, choose an arbitrary point in the dataset. and then check if there are minPts within eps distance around this point. Subsequently a new cluster is created if there are fewer than minPts within eps distance, the point is labeled as noise (which might later be found to be a border point to a cluster). If the point is a core point of another cluster, then the point could be part of the density border of a new cluster. The algorithm will connect these clusters together.

Subsequently, the method was employed to establish connections between these clusters. Each data field exhibits varying proportions of missing data. The absence of this component affects the structural model, resulting in a less precise or dependable outcome. The given data comprises four files that show the sales from December 2021 to the conclusion of April 2022.

Selecting the right features to construct the RFM is of utmost importance throughout the preprocessing step. The RFM Model necessitates the inclusion of specific qualities in order to construct an accurate RFM model. To analyze the RFM clusters effectively, we must first calculate the RFM quantitative values. This phase involved ranking products as customers and performed clustering on the data accordingly. The principle behind this process is to group products based on their last sales recency, frequency of sales, and total revenue generated. By doing so, we could understand which products were performing well and which may need some extra attention to increase sales and revenue. The common characteristics frequently utilized in RFM analysis encompass:

- Calculate Recency (R): For each product, we figured out the date of the most recent purchase. Then we subtracted this date from a snapshot date, which starts from the first of December 2021 until the end of May 2022. Then we analyzed and calculated the recency value, with lower values indicating more recent activity.\
- Calculate Frequency (F): We counted the sales made by each item. This would help represent the frequency value, with higher values indicating a higher frequency of purchases.
- Calculate Monetary Value (M): We added up the total value of all sales made by each customer. This would help define the monetary value, with higher values indicating greater total spending.

4. Analysis and Result

Two metrics, namely the Silhouette score and Davies-Bouldin score (Shahapure and Nicholas, 2020) were utilized in this study. These metrics are widely applied to evaluate the clustering findings' quality. The DBSCAN algorithm produced superior performance with a silhouette score of 0.890, which we obtained when the value of eps was set to 0.1 and the number of samples was ten. In contrast, the K-means algorithm earned a score of 0.724 with four clusters and a random state of twelve. On the other hand, the FCM yielded a silhouette score of 0.705, somewhat lower than the desired k-value of six and

fuzziness of 1.5. The Davies-Bouldin score (DBIndex) is an alternative assessment metric used to quantify the performance of unsupervised machine learning. The outcome as shown in Figure 2 demonstrates that DBSCAN attained a score of 1.607, which is in close proximity to the optimal value of 1.

On the other hand, the K-means algorithm and FCM obtained scores of 0.684 and 0.819, respectively. These scores were relatively low, suggesting that the clustering procedure effectively categorized the data according to the RFM principle. This suggests that the clusters are dense and distinct from each other. The K-Means algorithm successfully grouped the data points into clusters based on their sales trends, demonstrating good performance with 4 clusters as shown in Table 3.

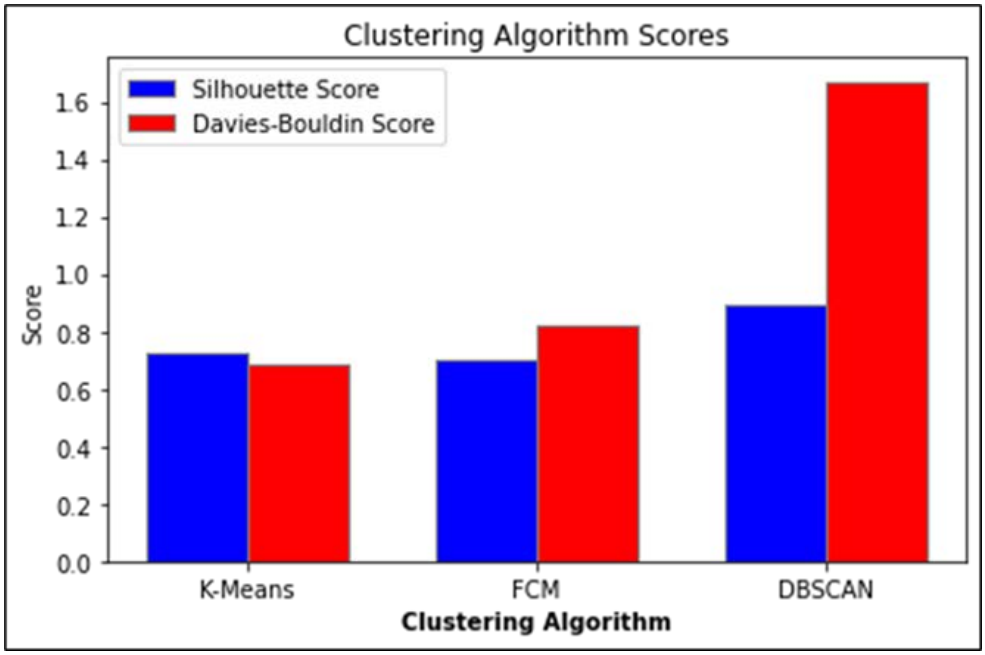

Figure 2: Clustering Algorithm Scores

Table 3: Performance measurement and the value

Clustering Algorithm	Parameters Setting	Accuracy
K-Means Algorithm	Number Of Cluster (K) = 4	Silhouette Score = 0.724 Davies-Bouldin score = 0.684
Fuzzy C-Means Algorithm (FCM)	Number Of Cluster (K) =6. Fuzziness =1.5	Silhouette Score = 0.705 Davies-Bouldin Score = 0.819
Density-Based Spatial Clustering of Applications with Noise (DBSCAN)	minpts =5, eps=0.25	Silhouette Score = 0.890 Davies-Bouldin Score = 1.607

The clustering model has successfully identified distinct product segments. The three RFM features proved crucial in segmenting customers based on their purchasing behavior, creating groups that varied significantly in their recency, frequency, and monetary value of purchases. Table 3 illustrates the performance measurement and the value that the model had achieved. The results show that DBSCAN algorithm performance was the highest with 0.89 for three clusters, while both the K-means and Fuzzy C-means algorithms achieved high scores also with 0.724 and 0.705, respectively.

It seems that within cluster three, there are several items that are quite popular and sell very well. To ensure that hypermarket don't run out of stock and disappoint their customers, it's important to have a solid inventory management system in place. By keeping track of the stock levels and replenishing items as needed, it can make sure that these high-demand items are always available for purchase. Additionally, running promotions and advertising these popular products can help to boost sales even further and maximize our profits. It's important to keep track of our stock levels, especially for cluster three where we have identified several popular items. How-ever, it's also important to identify slow-moving items in order to reduce the amount of stock we hold and free up cash that was tied up.

5. Conclusion and Future Recommendation

This study successfully conducted to gain a deep understanding of attributes that have influenced Hypermarket's sales in Muscat city. The ML analytics was employed in the hypermarket to execute product segmentation clustering. Specifically, unsupervised machine learning approaches were employed to construct models using three distinct algorithm types: the K-means algorithm for hard clustering, the Fuzzy C-means strategy for soft clustering, and the density-based clustering algorithm known as DBSCAN. The results indicate that the DBSCAN algorithm outperformed other algorithms in terms of Silhouette score and Davies-Bouldin score. Product clustering relies on several aspects that impact the process, including customer preferences, product attributes, geographical demands, culture, and time. Future research might prioritize the integration of real-time data to enable stakeholders to promptly react to shifts in customer behavior, therefore facilitating corporations to make real-time adjustments to their product offers. The inclusion of client identification can facilitate the implementation of a hyper clustering strategy, which aims to establish a connection between product and customer segmentation. This can lead to more personalized shopping experiences, increased customer satisfaction, and improved customer loyalty. Another suggestion is to expand the number of cities within the same region and analyze the trends of these cities in order to create tailored marketing campaigns and establish an efficient distribution plan.

Acknowledgments

We would like to thank Faculty of Artificial Intelligence, Universiti Teknologi Malaysia.

References

Alghanam, O. A., Al-Khatib, S. N., & Hiari, M. O. (2022). Data mining model for predicting customer purchase behavior in e-commerce context. International Journal of Advanced Computer Science and Applications, 13(2)

Alsabti, K., Ranka, S., & Singh, V. (1997). An efficient k-means clustering algorithm.

Alsahafi, M., Nour, M., Alhindi, A., & Farrash, M. (2022, December). Application of Unsupervised Learning and Data Clustering to Saudi Online Stores. In 2022 Fifth National Conference of Saudi Computers Colleges (NCCC) (pp. 71-78). IEEE.

Amos, C., Holmes, G.R. and Keneson, W.C. (2014), “A meta-analysis of consumer impulsive buying”, Journal of Retailing and Consumer Services, Vol. 21 No. 2, pp. 86-97

Bang, J., Cho, Y., & Kim, M. S. (2015). Getting business insights through clustering online behaviors. Modelling and Simulation in Engineering, 2015, 4-4

Brahmana, R. S., Mohammed, F. A., & Chairuang, K. (2020). Customer segmentation based on RFM model using K-means, K-medoids, and DBSCAN methods. Lontar Komput. J. Ilm. Teknol. Inf, 11(1), 32.

Brownlee, J. (2020). 10 Clustering Algorithms with Python. 10 Clustering Algorithms with Python.

Bruestle, S., Pappalardo, L., & Guidotti, R. (2019). Defining Geographic Markets from Probabilistic Clusters: A Machine Learning Algorithm Applied to Supermarket Scanner Data. Available at SSRN 3452058.

Chauhan, S., Banerjee, R., & Dagar, V. (2023). Analysis of impulse buying behaviour of consumer during COVID-19: An empirical study. Millennial Asia, 14(2), 278-299

Dobre, C., & Milovan-Ciuta, A. M. (2015). Personality influences on online stores customers behavior. Ecoforum Journal, 4(1), 9.

Drenik, G. 2021. 2020 Accelerated the Future of ECommerce. What Does That Means For 2021 (Online), Available: https://www.forbes.com/sites/garydrenik/2021/01/21/2020-accelerated-the-future-of-ecommerce-what-does-that-means-for-2021/?sh=372650387167 (Accessed Jan 21, 2021).

Erjavec, J., & Manfreda, A. (2022). Online shopping adoption during COVID-19 and social isolation: Extending the UTAUT model with herd behavior. Journal of Retailing and Consumer Services, 65, 102867

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd (Vol. 96, No. 34, pp. 226-231).

Fahim, A. M., Salem, A. M., Torkey, F. A., & Ramadan, M. (2006). An efficient enhanced k-means clustering algorithm. Journal of Zhejiang University-Science A, 7, 1626-1633.

Florez-Lopez, R., & Ramon-Jeronimo, J. M. (2009). Marketing segmentation through machine learning models: An approach based on customer relationship management and customer profitability accounting. Social Science Computer Review, 27(1), 96-117

Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on k-means. Indonesian Journal of Electrical Engineering and Computer Science, 18(1), 470-477.

Grimmeau, J. P. (2013). A forgotten anniversary: the first European hypermarkets open in Brussels in 1961. Brussels Studies. La revue scientifique pour les recherches sur Bruxelles/Het wetenschappelijk tijdschrift voor onderzoek over Brussel/The Journal of Research on Brussels.
Hengsberger, A. (2018). 4 reasons why innovations fail. LEAD Innovations Blog. Retrieved from https://www.lead-innovation.com/english-blog/why-innovations-fail

Hu, G., Zhang, Y., & Yang, Q. (2019, May). Transfer meets hybrid: A synthetic approach for cross-domain collaborative filtering with text. In The World Wide Web Conference (pp. 2822-2829)

Idowu, S., Annam, A., Rangaraja, E., & Kattukottai, S. (2019). Customer Segmentation Based on RFM Model Using K-Means, Hierarchical and Fuzzy C-Means Clustering Algorithms

Koofan, A. A. A., & Kaleem, M. (2019). Analyze and Enhance Sales in Lulu Supermarket using Data Mining Technology. Journal of Student Research.

Kseniia, K. (2018). Market Basket Visualization for Hypermarkets with the Use of Big Data Analytics.

Lee, D., & Hosanagar, K. (2019). How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. Information Systems Research, 30(1), 239-259.

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. International Journal of Research in Marketing, 37(3), 481-504

Mahfuza, R., Islam, N., Toyeb, M., Emon, M. A. F., Chowdhury, S. A., & Alam, M. G. R. (2022). LRFMV: An efficient customer segmentation model for superstores. Plos one, 17(12), e0279262.

Medrano, N., Olarte-Pascual, C., Pelegrín-Borondo, J., & Sierra-Murillo, Y. (2016). Consumer behavior in shopping streets: the importance of the salesperson's professional personal attention. Frontiers in psychology, 7, 125

Monalisa, S., & Kurnia, F. (2019). Analysis of DBSCAN and K-means algorithm for evaluating outlier on RFM model of customer behaviour. Telkomnika (Telecommunication Computing Electronics and Control), 17(1), 110-117.

Moon, S. K., Kumara, S. R., & Simpson, T. W. (2006, January). Data mining and fuzzy clustering to support product family design. In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Vol. 4255, pp. 317-325)

Nguyen, P. (2021). Influence of Recommender System Use on Consumer Decision Making (Master's thesis, Hanken School of Economics)

Shahapure, K. R., & Nicholas, C. (2020, October). Cluster quality analysis using silhouette score. In 2020 IEEE 7th international conference on data science and advanced analytics (DSAA) (pp. 747-748). IEEE

Tahiri, N., Mazoure, B., & Makarenkov, V. (2019). An intelligent shopping list based on the application of partitioning and machine learning algorithms. In proceedings of the 18th Python in Science Conference (SCIPY 2019)

Vehmas, K., Ervasti, M., Tihinen, M., & Mensonen, A. (2015). Digitalization boosting novel digital services for consumers. ACSIJ Advances in Computer Science: An International Journal, 4(4), 80-92.

Zhang, E., Li, H., Huang, Y., Hong, S., Zhao, L., & Ji, C. (2022). Practical multi-party private collaborative k-means clustering. Neurocomputing, 467, 256-265.

Zuo, Y., Yada, K., & Ali, A. S. (2016, December). Prediction of consumer purchasing in a grocery store using machine learning techniques. In 2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE) (pp. 18-25). IEEE.