IAIC International Conference Series
Vol. 4, No. 1, 2023, pp. 139 ~ 152
E-ISSN: 2774-5899 | P-ISSN: 2774-5880
DOI: https://doi.org/10.34306/conferenceseries.v4i1.645

Customer Segmentation: Transformation from Data to Marketing Strategy

[1*]Luciana Abednego, [2]Cecilia Esti Nugraheni, [3]Adelia Salsabina
[1,2,3]Informatics, Parahyangan Catholic University, Indonesia
E-mail: [1*]luciana@unpar.ac.id, [2]cheni@unpar.ac.id
***Corresponding author**

(Received November 8, 2023 Revised November 17, 2023 Accepted November 25, 2023, Available online December 19, 2023)

Abstract

Customer segmentation plays a crucial role in modern business strategies, enabling organizations to effectively target and personalize their marketing efforts and enhance customer relationships. Clustering algorithms have emerged as a powerful tool for segmenting customers based on their similarities and differences. We complement the data with an RFM model to support the clustering results. RFM, which stands for Recency, Frequency, and Monetary, is a model for segmenting customers based on their historical transaction data. This study aims to explore the concept of customer segmentation and the application of the RFM model combined with clustering algorithms in the real customer dataset of a company. It presents an overview of datasets, and introduces the RFM model and its components, emphasizing the significance of recency (how recently a customer made a purchase), frequency (how often a customer makes a purchase), and monetary value (the amount spent by a customer). It highlights the practicality of the RFM model in quantifying customer behavior and categorizing customers into distinct segments. It also explains popular clustering algorithms, analyzes experimental results, and concludes with future remarks on the potential of customer segmentation. We combine unsupervised (K-Means and DBSCAN clustering) and supervised machine learning methods to build customer clusters, label each cluster based on its characteristics, and propose a strategy for each cluster.

***Keywords**: Customer Segmentation, RFM Model, Clustering Algorithm, K-Means, DBSCAN*

1. Introduction

Customer segmentation involves dividing a heterogeneous customer base into distinct groups based on various characteristics such as demographics (age, gender, education, family composition, income), geographics (location, country, state), behavior (purchasing, media, and technology habits, participation and interest in activities), lifestyle (young achievers, striving singles, sustaining seniors), preferences, and purchase patterns [1]. Businesses can tailor their products, services, and marketing campaigns by identifying meaningful customer segments better to meet each group's specific needs and preferences. This leads to improved customer satisfaction, increased sales, and enhanced customer loyalty [2].

Clustering involves grouping objects that share common characteristics, and in this case, the aim is to create meaningful clusters for customer analysis [3]. Indicate K-Means as an effective algorithm for customer segmentation. K-Means is a prototype-based partition clustering technique that determines a specified number of clusters based on centroids [4].

2. Research Method

This section elaborates on the proposed methods and models of this study (see Figure. 1).

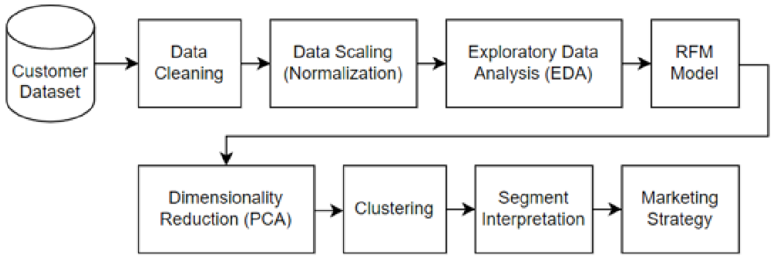

Figure. 1. Proposed Methodology

The process of customer segmentation and marketing strategy development involves several steps, i.e.:

1. Customer Dataset: We start with a dataset that contains information about customers of a company, which includes their demographics, purchase history, and any other relevant data [5]. This dataset serves as the foundation for analyzing and segmenting customers.
2. Data Cleaning: Data cleaning involves identifying and handling missing values, dealing with outliers, removing duplicates, and ensuring consistency in the dataset. This step helps improve the data quality and prevent any biases or inaccuracies in subsequent analyses [6].
3. Data Scaling: To avoid the dominance of certain features in subsequent analysis steps, we performed data scaling to ensure that the different features or variables in the dataset are on a similar scale.
4. Exploratory Data Analysis (EDA): To better understand the data's structure, we perform EDA, an iterative process that leads to additional questions and insights, prompting further exploration and refinement. It serves as a crucial foundation for the next clustering process [7].
5. RFM Model: RFM stands for Recency, Frequency, and Monetary Value. It is a customer segmentation model based on their purchasing behavior. In the RFM model, we assign a score to each customer based on how recently, how frequently they make purchases, and how much money they spend. These scores help to categorize customers into different segments, such as high-value, low-value, loyal, or churned customers [8].
6. Principal Component Analysis (PCA): PCA is a dimensionality reduction technique that helps to identify the most important patterns and relationships in the dataset. It reduces the dimensionality of the dataset by transforming the original variables into a new set of variables, called principal components. These components are ordered by their importance in explaining the variance in the data. PCA helps to simplify the dataset while retaining as much information as possible [9].
7. Clustering: After calculating the RFM scores, we apply clustering algorithms to group similar customers together. Clustering helps to identify distinct segments within the customer base. Each cluster represents a group of customers with similar characteristics and behaviors. We compare the performance of two clustering algorithms, namely k-Means and DBSCAN to build customer segmentation [10].
8. Segment Interpretation: After clustering, we analyze and interpret each customer segment to gain insights into their characteristics, preferences, and needs. This involves examining the RFM scores, demographic information, and any other relevant variables. We identify the key attributes that distinguish each segment and understand their potential value to the business [11].
9. Marketing Strategy: In future work, based on the segment interpretation, different marketing strategies can be developed for each customer segment. This involves creating targeted campaigns, personalized offers, and communication channels that are most effective for each segment. By understanding the unique needs and behaviors of each segment, marketing efforts can be optimized and enhance customer engagement and loyalty [12].

Overall, this process transforms raw customer data into meaningful segments to develop marketing strategies that resonate with each segment's preferences, ultimately leading to more effective customer targeting and improved business outcomes.

3. Results and Analysis

This section explains the results of the research and the comprehensive discussion.

3.1. Customer Dataset Collection

The proposed methodology is implemented on the real customer dataset of a credit guarantee company, P.T. Jamkrindo in Jambi City, Indonesia. P.T. Jamkrindo is the largest guarantee company in Indonesia. As a credit guarantee company, P.T. Jamkrindo has various products, both program guarantee products and non-program guarantee products [13]. In program guarantees, P.T. Jamkrindo has KUR guarantee products and KMK guarantees in the framework of PEN [14]. The dataset consists of 84,285 instances of customers with 37 attributes detailing the customer's habits and demographics, from January 2020 to December 2021. Missing values, incorrect types, inconsistent values, and outliers removal are handled using data preprocessing [15]. For the modified dataset, we apply RFM analysis and clustering algorithms to build customer segmentation or clusters. Table 1 shows all features for each customer record of the PT. Jamkrindo dataset.

Table 1. Customer Dataset

Column's Name	Column's Description
ID	Customer's unique identifier
Working Area	Working Area
City	The city where the office is located
Application Number	Application Number
Product	Product's name
Certificate Date	Date of the guarantee certificate
Bank/Financial institution	Bank/Financial institution used for disbursement
Unit Banks/Financial Institution	Unit Banks/Financial Institution
SK/SP	Guarantee certificate number
Name	Customer's name
Address	Customer Address
Birth Date	Customer's birth date
Age	Customer's age
ID Card	Customer's ID card number
NPWP	Customer's NPWP number
TK	TK value
Credit/Financing	Amount on credit taken
Guarantee/Kafalah	Amount of guarantee taken
IJP/IJK	Guarantee/Kafalah service fee
Project Value	Project value of this guarantee
Fee Base	Fee for base
Fee Agent	Fee for agency
Fire Compensation	Fire compensation value
Time Period	Period of guarantee taking
Time Unit	Unit of time for taking the guarantee
Contract Date	Date of the contract
Realization Date	Date of realization of the guarantee
Due Date	Due date of the guarantee

Column's Name	Column's Description
PK Number	Credit guarantee number at the credit realization
Instance	Instance used for guarantee
Allocation	Purpose of taking a guarantee
Account Number	Account number used for disbursement
Change Date	Date if there is any change
Working Area ID	ID for city where the office is located
Agent	Name of the agent used
Type of KUR	Type of KUR products selected
Reference Number	Guarantee application letter number

By analyzing this dataset, businesses can gain insights into customer demographics and purchase behavior. This information can be used to segment customers into distinct groups, develop personalized marketing strategies, target specific customer segments, and enhance customer satisfaction and loyalty [16]. Additionally, the dataset can be used to build predictive models to forecast customer behavior and identify potential high-value customers.

3.2. Data Preprocessing & Feature Engineering

After collecting the data, initial data preprocessing is carried out, which includes data cleaning, data transformation, data discretization, dealing with missing values, and replacing/dropping inappropriate values [17]. Some attributes are dropped since the correlation with the clustering segmentation is low. On the other hand, some new columns were added to apply the RFM model (see Table 2). As the last step in data preprocessing, the data is standardized so that all the features are in the same range.

Table 2. Some New Added Features

Column's Name	Column's Description
Year	The year of the transaction was carried out
Recency	Number of days since customer's last purchase, calculated from 1 January 2022
Frequency	Number of customer's purchases
Monetary	The total amount of customer's purchases

3.3. Exploratory Data Analysis

After data collection & preprocessing (data cleaning and scaling), the next step is to do an essential step in the data analysis process, which is called Exploratory Data Analysis (EDA) [18]. It involves the initial exploration and examination of a dataset to gain a deeper understanding of its structure, patterns, and characteristics. Before proceeding with more advanced analysis or modeling techniques, EDA is performed to uncover insights, identify relationships, and detect anomalies or outliers in the data [19].

As part of EDA, we try to analyze and understand customer behavior, preferences, and characteristics, which can be valuable for various business purposes, such as customer segmentation, personalized marketing, and customer relationship management [20].

3.3.1. Spending Patterns

In order to get insights into customers' spending habits and preferences for different product categories, we analyze attributes for different product categories [21]. This information can be used for product planning, cross-selling, and upselling strategies. Statistically, Figure 2 shows that:

1. Kredit Usaha Rakyat is the most significant product contributor to overall sales (94.3%).
2. 75.58% of customers choose Bank Rakyat Indonesia bank/financial institution as an intermediary between customers and P.T. Jamkrindo.
3. Two main allocation purposes are for working capital (68.36%) and investment (29.41%).

4. 99.99% of customers involve no agent.
5. The top three Kredit Usaha Rakyat (KUR) products are KUR Micro (67.37%), KUR Retail (18.19%), and KUR Super Micro (8.84%).
6. The most common period taken by the customer ranges from two to three years of guarantee.

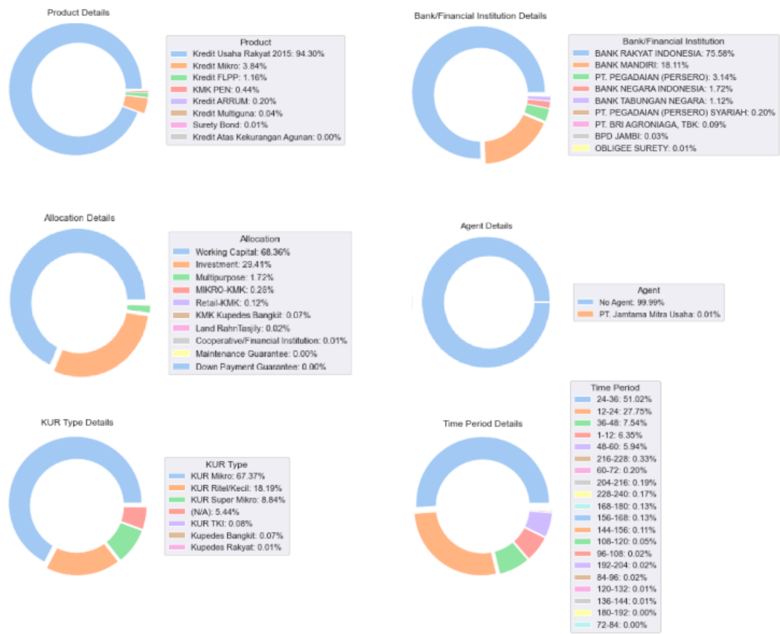

Figure 2. Exploratory Data Analysis of PT. Jamkrindo Customers Dataset

3.3.2. RFM Model

To analyze and segment customers based on their purchasing behavior, we use RFM model which stands for Recency, Frequency, and Monetary value. The RFM model evaluates customers based on three key metrics:

1. Recency: How recently a customer made a purchase
2. Frequency: How frequently a customer makes purchases
3. Monetary: How much a customer spends on purchases

RFM analysis assigns each customer a score for each of these metrics, which are then used to segment customers into groups based on their overall score [22]. This segmentation helps identify which customers are most valuable and should receive priority attention, and which customers are less valuable and may require less attention or targeted marketing efforts.

Figure 3 shows the violin plot that shows customer recency, frequency, and monetary value. Each violin plot shows the distribution of customer recency/ frequency/ monetary value for a particular category. The shape of the violin is created by mirroring a kernel density plot on each side of the median line. The violin's width corresponds to the density of customers at different levels. Wider sections indicate higher density, while narrower sections indicate lower density [23]. Violin plots can also help identify outliers or extreme values. If there are individual points or lines outside the violin shape, they represent customers with unusually high or low frequencies compared to the rest of the category.

Figure 3 shows a distribution of customer recency (the number of days since the last transaction). From Figure 3, we can conclude that the number of new transactions is greater than old ones, which means there was an increase in the number of customers. For frequency of purchases (Figure 4), most customers made 1 purchase, and a few customers made up to 16 purchases. Figure 5 shows that most transaction volume is below $3,000 (about IDR 37,000,000).

3.4. Clustering Algorithm

Clustering algorithms are machine learning techniques used to group similar data points together based on their characteristics or similarities [6]. These algorithms help in identifying inherent patterns, structures, or relationships within the data. Several clustering algorithms that we compared in this study:

3.4.1. K-means Clustering

K-means clustering is a widely used algorithm that partitions a dataset into a predefined number (k) of clusters. It starts by randomly selecting k cluster centroids and assigns each data point to the nearest centroid. The centroids are then updated by calculating the mean of the data points assigned to each cluster. This process is repeated iteratively until the centroids converge, and the clusters become stable. K-means clustering aims to minimize the within-cluster sum of squared distances, effectively grouping similar data points together. Algorithm 1 shows the pseudocode for the k-Means algorithm.

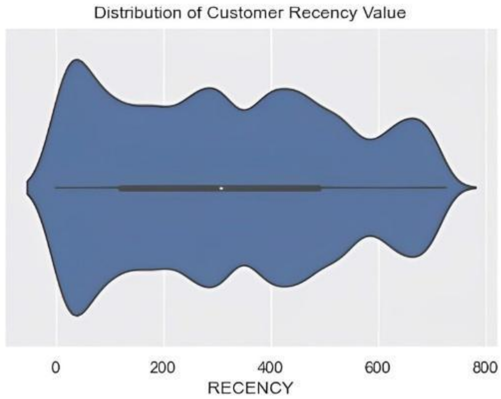

Figure 3. Distribution of Customer Recency

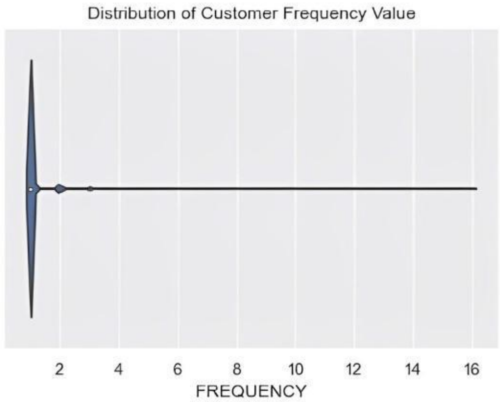

Figure 4. Distribution of Customers Frequency

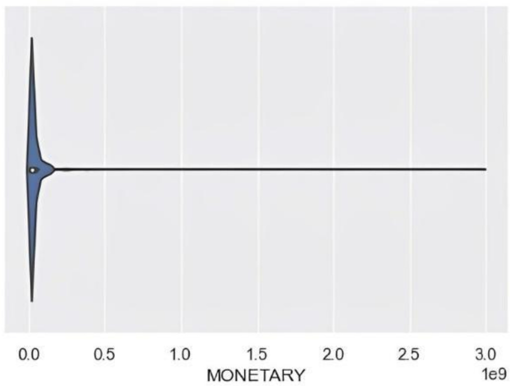

Figure 5. Distribution of Customers Monetary Value

Algorithm 1 K-Means Clustering Algorithm

```
Select k as the number of clusters
Randomly select k data points as centroids (assign as a cluster)
repeat
    for each data point do
        Calculate the distance of the data point to every centroid
        Assign the data point to the nearest centroid
        Recalculate the centroid as the mean of all data points in the cluster
    end for
until Convergence
```

Algorithm 1. K-Means Clustering Algorithm

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

DBSCAN is a density-based clustering algorithm that groups data points based on their density. It defines clusters as dense regions separated by sparser regions in the data space. The algorithm starts by selecting a random data point and expands the cluster by including neighboring points that have a sufficient number of nearby points (minPoints) within a specified distance threshold (epsilon). DBSCAN classifies points as core points, which have enough nearby points to form a cluster, border points, which are not core points but are within the epsilon distance of a core point, or noise or outlier points (Figure 6).

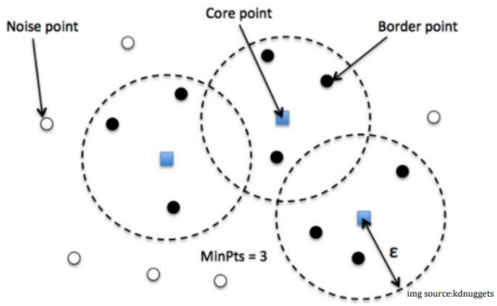

Figure 6. Noise, Core, and Border Points in DBSCAN Clustering

The DBSCAN algorithm can discover clusters of arbitrary shapes and is robust to noise. Figure 7 shows an illustration of DBSCAN compared to the k-Means clustering result. Algorithm 2 shows the pseudocode for the DBSCAN algorithm.

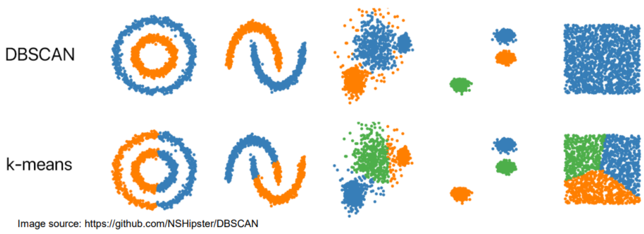

Figure 7. An Illustration of DBSCAN Clustering Result

3.5. Clustering Algorithm

We use the Elbow Method to determine the best number of clusters (k) for PT. Jamkrindo dataset. As the number of clusters (k) increases, the distortion score tends to decrease since each data point can be assigned to a closer centroid (Figure 8). However, beyond a certain point, the improvement in the distortion score becomes less significant, resulting in a flattened elbow shape in the plot. The elbow point represents the trade-off between the complexity (number of clusters) and the quality of the clustering.

Evaluating clustering algorithms helps determine the quality and effectiveness of the clustering results. We used two common internal evaluation methods for assessing clustering algorithms, i.e.: Silhouette Score and Davies Bouldin Index. The Silhouette Score is a measure of how well each data point in a clustering analysis fits within its assigned cluster, indicating the quality of the clustering results. It provides an estimate of the cohesion (how close a data point is to its own cluster) and separation (how far a data point is from other clusters) of the clusters. A higher Silhouette Score indicates better-defined and well-separated clusters.

Davies Bouldin Index quantifies the compactness and separation of clusters by considering both within-cluster scatter and between-cluster distances. It measures the average dissimilarity between each cluster and its most similar neighboring cluster, taking into account both the within-cluster scatter and the between-cluster separation. The DBI provides a single numerical value, where a lower value indicates better-defined and well-separated clusters. Based on Table 3:

1. DBSCAN has the highest Silhouette Index, indicating better clustering results in terms of cluster cohesion and separation compared to k-Means.
2. DBSCAN has the lowest Davies Bouldin Index, suggesting better clustering results in terms of cluster compactness and separation compared to k-Means.

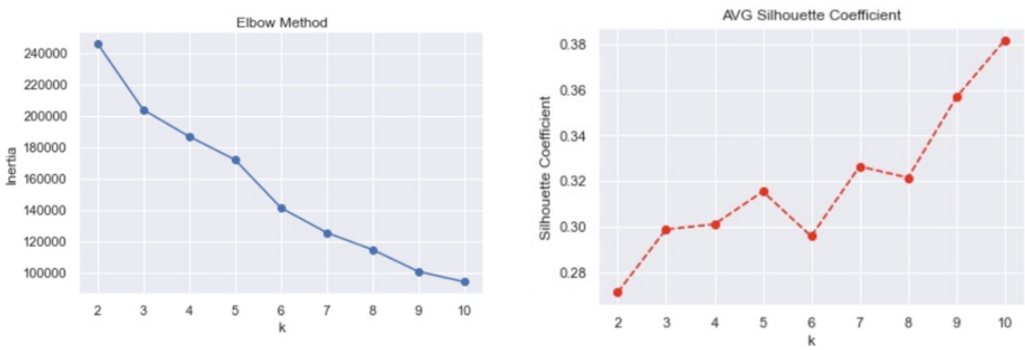

Figure 8. Elbow Method in k-Means & Its Silhouette Score

Algorithm 2 DBSCAN Clustering Algorithm

```
Mark all data points as unvisited
repeat
    Randomly select an unvisited data point m
    Mark m as visited
    if m is a core point (it has at least minPoints in eps-neighborhood) then
        Create a new cluster A, and add m to cluster A
        Let N be the set of data points in the ε-neighborhood of m
        for each data point n in N do
            if n is unvisited then
                Mark n as visited
                if n is a core point then
                    Add n to N
                end if
            end if
            if n is not yet a member of any cluster then
                Add n to cluster A
            end if
        end for
    else
        Mark m as noise (because it's not assigned to any cluster and not a
core point)
    end if
until All data points are visited
```

Algorithm 2. DBSCAN Clustering Algorithm

Table 3. Evaluation Metrics Of K-Means, Agglomerative, And Dbscan Clustering

Algorithm	Silhouette Index	Davies Bouldin Index
k-Means	0.2996	1.19
DBSCAN	1.19	1.00

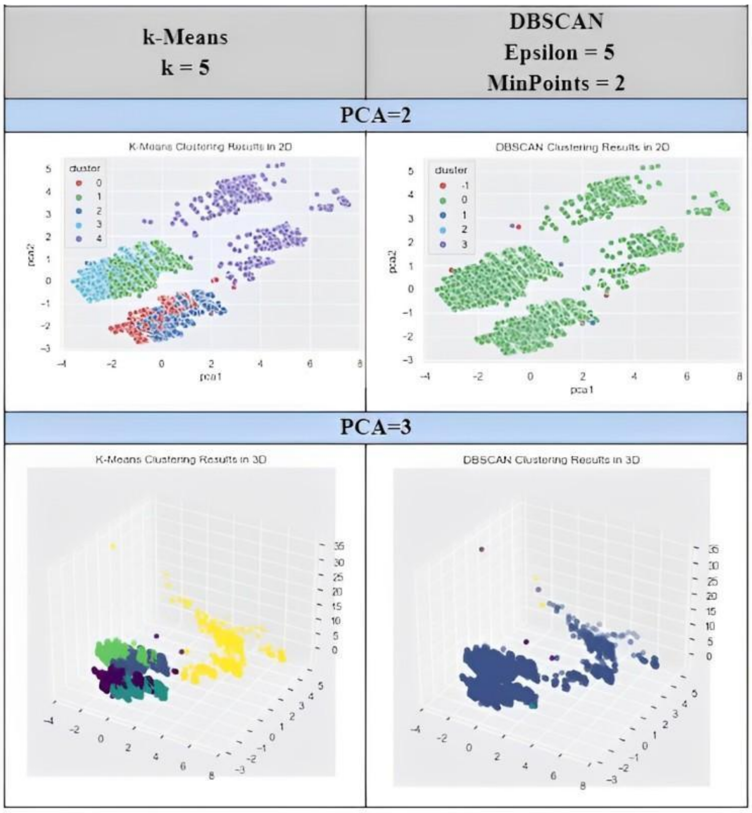

Figure 9. Distribution of Number of Customers Per Cluster

3.5. Segment Interpretation And Marketing Strategy

After the clusters have been generated, the next step is interpreting and extracting meaningful insights from these clusters. We examine the characteristics of each cluster, such as the mean or median values of the features within the cluster, look for some cluster characteristics with high within-cluster similarity and low between-cluster similarity, and identify the distinguishing features or patterns that differentiate one cluster from another. Figure 10 – Figure 12 show the distribution of customer recency, frequency, and monetary.

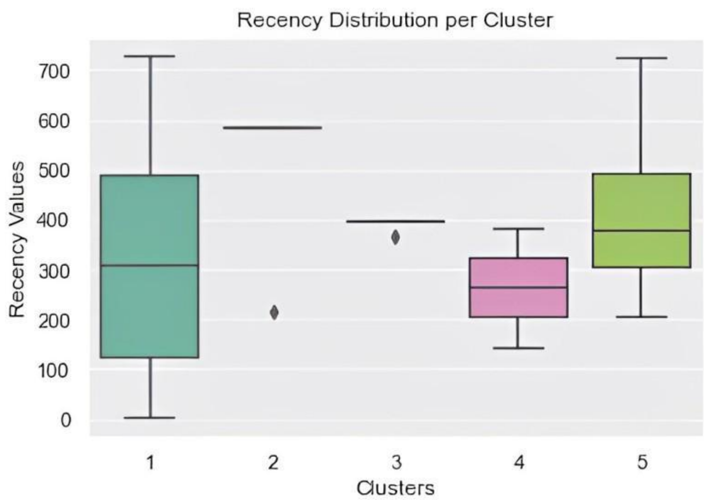

Figure 10. Distribution of Customers Recency

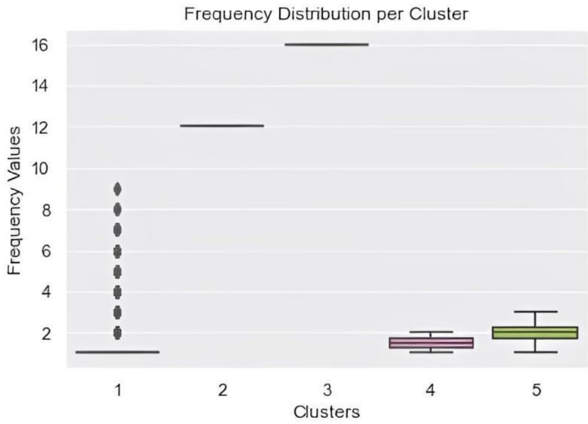

Figure 11. Distribution of Customers Frequency

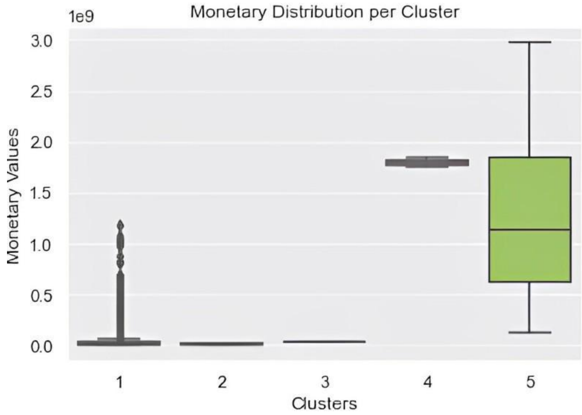

Figure 12. Distribution of Customers Monetary

We combine the unsupervised method to build clusters of customers and the supervised method to label each cluster with some categories. Table IV-VI shows some category labels with their boundaries. Based on the label as stated in Table 4-6, the interpretation for each segment or cluster is as Table 7-11.

Table 4. Label's Category For Recency

Label	Description
Latest	Clusters that have the newest recency: 263 to 349 days since the last purchase. Calculated from 1 January 2022.
Middle	Cluster with 350 to 434 days since the last purchase
Longest	Cluster with 435 to 520 days since the last purchase

Table 5. Label's Category For Frequency

Label	Description
Low	Cluster with 1 to 6 number of transactions
Medium	Cluster with 7 to 11 number of transactions
High	Cluster with 12 to 16 number of transactions

Table 6. Label's Category For Monetary

Label	Description
Low	Cluster with a mean of monetary IDR 15,138,668 to IDR 609,659,129.
Medium	Cluster with a mean of monetary IDR 609,659,130 to IDR 1,204,179,591.
High	Cluster with a mean of monetary IDR 1,204,179,592 to IDR 1,798,700,050

Table 7. Cluster 1 Profile & Strategy

Cluster's Name	Cluster 1 - Main Customers
Description	The most members of the company
Recency	Medium
Frequency	Low
Monetary	Low
Maximum term	20 years
Strategy	Create a loyalty program to keep them purchasing on a regular basis. Since it has the most members, it should have the most influence on the company.

Table 8. Cluster 2 Profile & Strategy

Cluster's Name	Cluster 2 – Dormant Customers
Description	Customers who are likely to churn. The last purchase was long back, low spenders, but bought often.
Recency	Longest
Frequency	High
Monetary	Low
Maximum term	2-3 years
Strategy	Offer exclusive promotions or discounts to dormant customers as an incentive to return.

Table 9. Cluster 3 Profile & Strategy

Cluster's Name	Cluster 3 – Promising Customers
Description	Customers who have not made purchases recently, haven't spent much, but often.
Recency	Middle
Frequency	High
Monetary	Low
Maximum term	3-4 years
Strategy	Retain them with exciting offers. Make offers to bring them back to purchasing

Table 10. Cluster 4 Profile & Strategy

Cluster's Name	Cluster 4 – Potential New Customer
Description	Customers who have made purchases recently, who spent a good amount, but not often.
Recency	Latest

Cluster's Name	Cluster 4 – Potential New Customer
Frequency	Low
Monetary	High
Maximum term	2 years
Strategy	Create a loyalty program so that they can continue spending more.

Table 11. Cluster 5 Profile & Strategy

Cluster's Name	Cluster 5 - Potential Customers
Description	At Risk: Spent big money but a quite long time ago. Need to bring them back.
Recency	Middle
Frequency	Low
Monetary	High
Maximum term	3 years
Strategy	Send them an interesting new offer

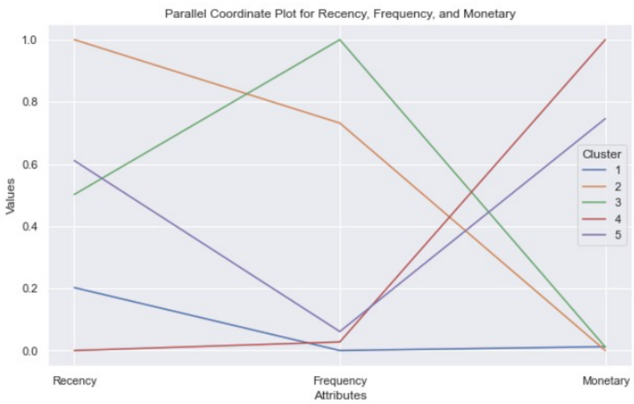

Figure 13. Recency, Frequency, and Monetary for each Cluster

Figure 13 shows the parallel coordinate visualization for all clusters that are mapped to Recency, Frequency, and Monetary values. Clusters 2 and 3 have tendencies of high frequencies but low monetary value. On the other hand, Clusters 4 and 5 have low frequencies but high monetary value. Cluster 1 is new customers with low frequency and monetary value.

4. Conclusion

In this research, we combine unsupervised and supervised machine learning methods to segment customers of a company and interpret the result. We compare two clustering techniques, k-Means and DBSCAN to build clusters of customers. From the experiment results, DBSCAN outperforms k-Means algorithm since it has a higher silhouette score and a lower Davies Bouldin Index. DBSCAN clusters the dataset so that its intra-cluster data points are very similar to each other than other data points in other clusters.

K-means and DBSCAN are suited to different types of data and clustering scenarios. The choice between the two depends on the nature of the data and the specific goals of clustering analysis. K-means is appropriate when you have prior knowledge of the number of clusters and when clusters are compact and equally sized. On the other hand, DBSCAN is suitable for cases where the number of clusters is unknown, clusters have irregular shapes, and noise or outliers need to be identified.

The customer cluster is then labeled based on its features and some marketing strategies are suggested (Table 4-11 and Figure 13). From this result, the company can give different approaches for each customer cluster, and optimize their marketing strategies, product development, and customer service efforts. It enhances customer engagement, loyalty, and revenue while helping companies adapt to evolving customer preferences and market dynamics.

Several potential directions for future research are to take into account several modified parameters with the RFM model, like customer satisfaction [4] or duration between the first and the last purchase of a customer [7].

References

[1] W. Qadadeh and S. Abdallah, "Customers Segmentation in the Insurance Company (TIC) Dataset", Procedia Computer Science Volume 144, 2018, Pages 277-290, 2018.
[2] R. Widayanti, M. H. R. Chakim, C. Lukita, U. Rahardja, and N. Lutfiani, "Improving Recommender Systems using Hybrid Techniques of Collaborative Filtering and Content-Based Filtering," J. Appl. Data Sci., vol. 4, no. 3, pp. 289–302, 2023.
[3] M.-H. Lin et al., "Capacity Optimization Design of Hybrid Energy Power Generation System.," Math. Model. Eng. Probl., vol. 10, no. 4, 2023.
[4] U. Rahardja, Q. Aini, D. Manongga, I. Sembiring, and I. D. Girinzio, "Implementation of Tensor Flow in Air Quality Monitoring Based on Artificial Intelligence," Int. J. Artif. Intell. Res., vol. 6, no. 1, 2023.
[5] M. Mosa et.al., "A Novel Hybrid Segmentation Approach for Decision Support: A Case Study in Banking", The Computer Journal, Volume 66, Issue 5, May 2023, Pages 1228-1240, 2022.Trans. Sustain. Digit. Innov., vol. 4, no. 2, pp. 128–136, 2023.
[6] A. G. Prawiyogi, "Southeast Asia's Cyber Security Strategy: Multilateralism or Self-help," IAIC TA. Dudhat and V. Agarwal, "Indonesia's Digital Economy's Development," IAIC Trans. Sustain. Digit. Innov., vol. 4, no. 2, pp. 109–118, 2023.
[7] M. F. Fazri, L. B. Kusuma, R. B. Rahmawan, H. N. Fauji, and C. Camille, "Implementing Artificial Intelligence to Reduce Marine Ecosystem Pollution," IAIC Trans. Sustain. Digit. Innov., vol. 4, no. 2, pp. 101–108, 2023.
[8] A. Cahyono and Y. D. Nurcahyanie, "Identification and Evaluation of Logistics Operational Risk Using the Fmea Method at PT. XZY," Aptisi Trans. Technopreneursh., vol. 5, no. 1Sp, pp. 1–10, 2023.
[9] R. D. Affandi, H. Pratiwi, and M. I. Sa'ad, "Application of the SMARTER Method in Determining the Whitening of Study Permits and Teacher Study Tasks," Aptisi Trans. Technopreneursh., vol. 5, no. 2, pp. 205–215, 2023.
[10] M. Miran and O. Sumampouw, "Superior College Applied Research Competence of SPI Members in the Context of Improving the Quality of Supervisory Performance at Manado State University," Aptisi Trans. Technopreneursh., vol. 5, no. 1, pp. 73–86, 2023
[11] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," IEEE Access, vol. 8, pp. 80716-80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
[12] O. A. D. Wulandari and D. Apriani, "Sustainable Institutional Entrepreneurial Culture and Innovation For Economic Growth," APTISI Trans. Manag., vol. 7, no. 3, pp. 221–230, 2023.
[13] [48] E. Nathanael and W. Sejati, "Effective Government Management of Flood Discharge in Drainage Channels using HEC-RAS 6.3. 1 Application," APTISI Trans. Manag., vol. 7, no. 3, pp. 210–220, 2023.
[14] R. Rarmizi, I. Y. Nasaruddin, and N. Hidayah, "Analysis of the influence of corporate governance on the financial performance of Islamic banks in Indonesia 2016-2021," APTISI Trans. Manag., vol. 7, no. 2, pp. 179–190, 2023.
[15] D. Mensouri et.al., "K-Means Customers Clustering by their RFMT and Score Satisfaction Analysis", International Journal of Advanced Computer Science and Applications, Vol. 13, No. 6, 2022.
[16] D. Iriani, S. Parman, A. F. Hafizh, I. Rachmawati, and Y. A. Solihah, "Ambient Media Advertisement of Catur Insan Cendekia University to Improve Brand Awareness," ADI J. Recent Innov., vol. 5, no. 1Sp, pp. 97–110, 2023.
[17] C. Lukita, M. H. R. Chakim, R. Supriati, N. P. L. Santoso, and M. F. Kamil, "Exploration of Perceived Use of Technology Using A Digital Business Perspective," ADI J. Recent Innov., vol. 5, no. 1Sp, pp. 87–96, 2023.
[18] F. P. Oganda, M. H. R. Chakim, W. E. Septian, and E. D. Astuti, "User Involvement on Air Quality in Incubation Rooms in Banten-Indonesia," ADI J. Recent Innov., vol. 5, no. 1, pp. 86–92, 2023. [7] M. Annas and S. N. Wahab, "Data Mining Methods: K-Means Clustering Algorithms," Int. J. Cyber IT Serv. Manag., vol. 3, no. 1, pp. 40–47, 2023.

[19] [8] U. Usanto, L. Nurlaela, A. Sopian, and F. Alfiah, “Umrah Registration System Using Extreme Programming Method Towards Worship Tourism,” Int. J. Cyber IT Serv. Manag., vol. 3, no. 1, pp. 22–31, 2023.

[20] [9] E. Nurninawati, M. Y. Effendy, and A. M. Rianputra, “Web-Based Product Marketing Information System Design at Definier Store,” Int. J. Cyber IT Serv. Manag., vol. 3, no. 1, pp. 1–11, 2023.

[21] A. Zheng, A. Casari, "Feature Engineering for Machine Learning_ Principles and Techniques for Data Scientists", O'Reilly Media, 2018.

[22] C. C. Aggarwal, C. K. Reddy, "Data Clustering Algorithms and Applications", Taylor & Francis Group, LLC. 2014.

[23] M. Alvandi, S. Fazli, and F. Abdoli, “K-Mean Clustering Method For Analysis Customer Lifetime Value With LRFM Relationship Model In Banking Services,” 2012. https://www.semanticscholar.org/paper/KMean-Clustering-Method-For-Analysis-Customer-With-AlvandiFazli/54ac58708a608bee195e9f60825a61da51ebbcf1 (accessed October 20, 2023).