International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 04 April 2025, pp: 34-43 www.ijaem.net ISSN: 2395-5252

Comparative Analysis of Clustering Techniques for Customer Segmentation: Evaluating K-Means, Hierarchical, and DBSCAN Models alongside RFM Frameworks to Enhance Marketing Strategies through Behavioral, Demographic, and Transactional Insights

Titus Zira Fate, Dogo Siyani Ezra & Ijandir Isaac Samuel
Nigerian Army University Biu Department of Information Technology
Taraba State University Department of Computer Science
The American University of Nigeria Department of Information Systems

Date of Submission: 25-03-2025 Date of Acceptance: 05-04-2025

ABSTRACT
The study aims at clustering, segmentation customer using K-means clustering model, hierarchical clustering model, Density-based Spatial Clustering of Applications with Noise (DBSCAN) model and customer segmentation frameworks (RFM), A few unsupervised machine learning (ML) clustering models such as K-means clustering model, hierarchical clustering model, Density-based Spatial Clustering of Applications with Noise (DBSCAN) and customer segmentation frameworks (RFM), to identify distinct and actionable customer segment based on their behavioral, demographic and transactional characteristics. The traditional model was included in the analysis since clustering models are not optimization models and the goodness of unsupervised models could only be evaluated with a practical business approach. Theresults and discussion of the paper highlightscustomer segmentation based on their historical transactional characteristics, evaluating the effectiveness of different clustering algorithms and customer segmentation framework the potential for future enhancements. The emphasis on statistical analysis and the evaluation of various clustering techniques provide valuable insights into effective customer segmentation strategies

Keywords: Machine Learning, Customer Clustering, & Segmentation

I. INTRODUCTION

Machine learning-based customer clustering and segmentation is a vital strategy for businesses aiming to enhance customer engagement and tailor marketing efforts. By employing various algorithms, such as K-means, hierarchical clustering, and DBSCAN, organizations can identify distinct customer segments based on shared characteristics, leading to more effective marketing strategies and improved customer satisfaction. While machine learning offers powerful tools for customer segmentation, it is essential to recognize that the effectiveness of these methods can vary based on data quality and the chosen algorithms. Businesses must continuously refine their approaches to adapt to changing customer behaviors and preferences. Therefore, the use of data mining techniques is one of the ways to solve the problem of customer segmentation for developing new market strategy. Data mining is extremely powerful tool and has change many companies fortune, Identification of customers is based on data collect from customers and grouping them in some meaningful order(Shirole, Salokhe, &Jadhav, 2021)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 04 April 2025, pp: 34-43 www.ijaem.net ISSN: 2395-5252

Hierarchical Clustering is an analytical approach constructs, a dendrogram of clusters, facilitating a more intricate comprehension of customer interrelations(Rahmet & Gürkan, 2024). Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is the density-based clustering algorithm discerns clusters of diverse geometric shapes and sizes, rendering it particularly adept for the analysis of complex datasets(Preemi, Sriram,Deva, Suji, &Reddy 2024).Conversely, although the RFM model and K-means clustering model demonstrates efficacy, certain studies propose that alternative clustering methodologies, such as Hierarchical clustering, may facilitate more precise segmentations due to their superior capacity to accommodate data distributions and inter-variable relationships (Wong, Ying,Yin,Yi, Xiao, Duan, He, &Tang 2024). Data clustering is divided into two purposes referring namely clustering for understanding and clustering for use. If the goal is to capture the natural structure of the data (He & Tan, 2012). Usually the clustering process is then followed by the summarization process mean, standard deviation class labeling in each group to be used as classification training data and so on.

The market segmentation is a process to divide customers into homogeneous group which have similar characteristics such as buying habits, life style, food preference. Data mining serves as a crucial tool for revealing latent patterns within consumer data, thereby enhancing the efficacy of customer segmentation. Customer segmentation involves the categorization of analogous consumers, which assists enterprises in directing their marketing initiatives and elevating customer satisfaction. This methodological approach empowers organizations to tailor marketing strategies, extend targeted discounts, and gain deeper insights into customer relationships. Ultimately, proficient segmentation has the potential to augment revenue by more accurately identifying and addressing the needs of target customers(A et al., 2024). The introduction of Personalized Marketing Importance underscores the escalating relevance of personalized marketing within the contemporary competitive business milieu. Organizations are increasingly acknowledging that customized offers and distinctive promotions can substantially augment customer engagement and foster loyalty. Data-Driven Strategies delineates the transition towards data-driven methodologies, particularly the application of machine learning techniques, to more effectively comprehend and address customers' specific requirements and preferences. This paradigm is indispensable for enterprises striving to flourish in a saturated marketplace.The papers investigate analysis of various techniques which is presented on customer segmentation methods based on online retails data. The traditional model is also included in the analysis since clustering model are not optimized models

Statement of the Problem

The study aims at clustering, segmentation customer using K-means clustering model, hierarchical clustering model, Density-based Spatial Clustering of Applications with Noise (DBSCAN) model and customer segmentation frameworks (RFM). The model will analyze customer data, such as demographic information, purchase history, and behavioral data, to identify patterns and similarities between customers. The customers will then be grouped into distinct clusters based on these similarities. The goal is to use these clusters to better understand the needs and preferences of each group, and to develop targeted marketing and service strategies for each group. The model will be evaluated based on its ability to accurately cluster customers.

Research Objectives

i. To analyzed customers clustering segmentation using K-means clustering model, hierarchical clustering model, Density-based Spatial Clustering of Applications with Noise (DBSCAN) model and customer segmentation frameworks (RFM),
ii. To analyze the customer data such as demographic information, purchase history, preferences and behavioral data
iii. To evaluates various unsupervised machine learning clustering models

II. LITERATURE REVIEW

The proliferation of big data and the growth of e-commerce have intensified the challenges associated with extracting actionable data for personalized(Wong et al., 2024).Shen propose a framework utilizing sentiment networks derived from online reviews, employing graph embedding techniques to represent customer sentiments as vectors. This method enhances segmentation accuracy by capturing nuanced customer opinions, leading to improved product design and marketing strategiesShen, Feng, Cheng,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 04 April 2025, pp: 34-43 www.ijaem.net ISSN: 2395-5252

& Bi, 2024).In a related report (Gupta & Israni, 2024)illustrate the capacity of machine learning to yield significant insights derived from customer data, thereby augmenting organizational performance and fostering customer engagement.Studies have been carried out by (Ullah, Mohmand,Hussain, Johar, Khan, Ahmad, Mahmoud, et a&Hudal, 2023). Which was termed RFMT, the paper systematically examined the preeminent e-commerce dataset within Pakistan by integrating k-means, Gaussian, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) alongside agglomerative algorithms for the purpose of segmentation, culminating in the identification of three distinct clusters. A multitude of machine learning algorithms has been employed for the purpose of forecasting customer attrition, it emphasized the identification of the most efficacious approaches and the comprehension of the determinant that influence attrition. Though the model uses prediction of customer attrition within the banking industry through the application of machine learning classification methodologies (Nguyen, Le,Dao, & T, 2023). Research has been done byHicham & Karim, (2022) aim to introduce a novel customer segmentation technique that enhances clustering quality by integrating multiple clustering results through a clustering ensemble and spectral clustering. Different clustering algorithms such as K-Means, Agglomerative and Mean Shift) are been implemented to segment the customer which are finally compared with the results obtained from the algorithms (Vardhan & Sharma, 2023), furthermore (Huang, 2024)focuses on customer segmentation using the Rubin Index based K means clustering approach, addressing challenges in partitioning customers effectively. It emphasizes data quality improvement and evaluates model performance through metrics like accuracy, precision, and F1-score. (Nugroho, Rafhina, Ananda, & Gunawan, 2024)focuses on customer segmentation using K-Means clustering, highlighting its effectiveness in grouping customers based on sales transaction data. It contributes to literature by demonstrating optimized clustering techniques for targeted marketing strategies and enhancing customer data analysis.

Clustering Algorithms used in the research

The relevance of using customer clustering and segmentation using machine learning is a vital strategy for businesses aiming to enhance customer engagement and optimize marketing efforts. Various algorithms, particularly K-means, hierarchical clustering, and DBSCAN, have been employed to analyze customer data effectively. These methods allow businesses to identify distinct customer segments based on shared attributes, leading to tailored marketing strategies and improved customer satisfaction.

i. K-Means Clustering:This unsupervised algorithm groups customers by minimizing the distance between data points and cluster centroids, effectively identifying distinct segments (Loganathan & Reddy, 2024)
ii. Gaussian Mixture Models (GMMs) are versatile probabilistic models that effectively represent data generated from multiple Gaussian distributions. They utilize a soft clustering approach, assigning probabilities to data points for each cluster rather than definitive assignments. The Expectation-Maximization (EM) algorithm is commonly employed to iteratively estimate the parameters of these distributions and the associated probabilities. This method has been shown to outperform traditional density approximation techniques in various applications, including neural networks and astronomical data analysis (López-Lobato & Avendaño Garrido, 2023).
iii. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a powerful clustering algorithm that excels in identifying clusters of varying shapes and densities, making it particularly useful in diverse applications such as disease outbreak analysis and crime data clustering. This algorithm operates by grouping together points that are closely packed together while marking points in low-density regions as noise. The following sections elaborate on its applications, enhancements, and limitations(Simbolon & Friskila, 2024)
iv. RFM is a customer segmentation technique used in marketing to identify and group customers based on their purchasing behavior. It involves analyzing three factors: Regency (how recently a customer made a purchase), frequency (how often a customer makes purchases), and monetary (how much money a customer spends). The data is then clustered into groups based on these factors to identify different customer segments with different characteristics and behavior. (Namvar, Ghazanfari , & Naderpour 2017)

III. METHODOLOGY

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 04 April 2025, pp: 34-43 www.ijaem.net ISSN: 2395-5252

Data Collection:

Data was collected online through secondary source which was obtained from customer online retails from Kaggle (https://www.kaggle.com/datasets/). The dataset contains information about comprehensive analysis of a company's preferred customers which enables businesses to gain a better understanding of their customers and customize their products to meet the specific needs, behaviors, and concerns of different customer segments and includes demographic information, product information, and responses to marketing campaigns. By analyzing the different customer segments, businesses can identify which segment is more likely to purchase a particular product. Thus which allows them to market the product effectively to that segment instead of spending money on marketing to all customers in the database. Below is the attached sample of the data used with the following variable such as Invoice No, Stock code, Description, Quantity, Invoice Date, Unit Price, Customer ID and Country.

Data Set/ Source of Data from https://www.kaggle.com/datasets/).

Data Preprocessing: Data was cleaning using handle missing values, duplications was removed, data types was corrected, negative or zero values in quantity and unit price. Data was transformed, splitting then apply data formatting, standardization.

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536365	85123A	WHITE HANGING H	6	12/1/2010 8:26	2.55	17850	United Kingdom
536365	71053	WHITE METAL LANT	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84406B	CREAM CUPID HEAI	8	12/1/2010 8:26	2.75	17850	United Kingdom
536365	84029G	KNITTED UNION FL	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	84029E	RED WOOLLY HOTT	6	12/1/2010 8:26	3.39	17850	United Kingdom
536365	22752	SET 7 BABUSHKA N	2	12/1/2010 8:26	7.65	17850	United Kingdom
536365	21730	GLASS STAR FROST	6	12/1/2010 8:26	4.25	17850	United Kingdom
536366	22633	HAND WARMER UN	6	12/1/2010 8:28	1.85	17850	United Kingdom
536366	22632	HAND WARMER RE	6	12/1/2010 8:28	1.85	17850	United Kingdom
536367	84879	ASSORTED COLOUF	32	12/1/2010 8:34	1.69	13047	United Kingdom
536367	22745	POPPY'S PLAYHOUS	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22748	POPPY'S PLAYHOUS	6	12/1/2010 8:34	2.1	13047	United Kingdom
536367	22749	FELTCRAFT PRINCE	8	12/1/2010 8:34	3.75	13047	United Kingdom
536367	22310	IVORY KNITTED MU	6	12/1/2010 8:34	1.65	13047	United Kingdom
536367	84969	BOX OF 6 ASSORTE	6	12/1/2010 8:34	4.25	13047	United Kingdom
536367	22623	BOX OF VINTAGE JI	3	12/1/2010 8:34	4.95	13047	United Kingdom
536367	22622	BOX OF VINTAGE A	2	12/1/2010 8:34	9.95	13047	United Kingdom
536367	21754	HOME BUILDING BL	3	12/1/2010 8:34	5.95	13047	United Kingdom
536367	21755	LOVE BUILDING BLC	3	12/1/2010 8:34	5.95	13047	United Kingdom
536367	21777	RECIPE BOX WITH N	4	12/1/2010 8:34	7.95	13047	United Kingdom
536367	48187	DOORMAT NEW EN	4	12/1/2010 8:34	7.95	13047	United Kingdom
536368	22960	JAM MAKING SET V	6	12/1/2010 8:34	4.25	13047	United Kingdom

Method Used in Data Analysis in Customer Clustering

Data Analysis Techniques:

The primary method used for analyzing customer data included demographic information, purchase history, and behavioral data. This comprehensive approach allows for a deeper understanding of customer segments and their specific need.

Machine Learning Models:
The research evaluated various unsupervised machine learning clustering models, including:
K-Means Clustering: A popular method for partitioning data into distinct clusters based on similarity.
Hierarchical Clustering: This method builds a hierarchy of clusters, allowing for a more detailed analysis of customer relationships.
Density-based Spatial Clustering of Applications with Noise (DBSCAN): This model identifies clusters based on the density of data points, making it effective for discovering clusters of varying shapes and sizes.

Traditional RFM Model:
The study also included a traditional model based on recency, frequency and monetary (RFM) analysis, which provides insights into customer behavior from a business perspective.

Statistical Analysis:
The study emphasizes the importance of statistical analysis in achieving business goals, particularly in understanding customer behavior and preferences. This analysis is crucial for tailoring marketing strategies and improving customer satisfaction.

Dimensionality Reduction
To aid in the visualization of the pre-processed dataset, Principal Component Analysis (PCA) was applied to reduce the dimensionality of the feature space. PCA is a commonly used technique for dimensionality reduction that transforms the original features into a new set of orthogonal features called principal components. These new features are ranked based on the amount of variance they explain in the original dataset, with the first principal component explaining the most variance and subsequent components explaining progressively less. It is important to note that in this studies, PCA was only used for visualization purposes and not for feature selection.

Implementation, Testing and Results
The process of developing a system entails developing or purchasing the necessary components, then integrating them to produce the finished system. The conceptual design and the development phases are connected by the system requirements. Business needs as well as high-level functional and non-functional system requirements are established during conceptual design.During the development phase there are activities carried out which are requirement engineering, Architectural design, requirement partitioning, systems integration, system testing, system development, unit and integration testing.

```
[2]: #load dataset
customer_df = pd.read_excel('OnlineRetail.xlsx')
customer_df.head(10)
```

[2]:

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
5	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	2010-12-01 08:26:00	7.65	17850.0	United Kingdom
6	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	2010-12-01 08:26:00	4.25	17850.0	United Kingdom
7	536366	22633	HAND WARMER UNION JACK	6	2010-12-01 08:28:00	1.85	17850.0	United Kingdom
8	536366	22632	HAND WARMER RED POLKA DOT	6	2010-12-01 08:28:00	1.85	17850.0	United Kingdom
9	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	2010-12-01 08:34:00	1.69	13047.0	United Kingdom

Figure 2 Loading Dataset and displaying information about first ten loaded dataset

```
[3]: customer_df.info()
print(customer_df.shape)

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    541909 non-null  object
 1   StockCode    541909 non-null  object
 2   Description  540455 non-null  object
 3   Quantity     541909 non-null  int64
 4   InvoiceDate  541909 non-null  datetime64[ns]
 5   UnitPrice    541909 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      541909 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 33.1+ MB
(541909, 8)
```

Figure 3 Getting Duplicate Value from dataset

```
[5]: print(customer_df.isnull().sum())

InvoiceNo           0
StockCode           0
Description      1454
Quantity            0
InvoiceDate         0
UnitPrice           0
CustomerID     135080
Country             0
dtype: int64
```

Figure 4. Getting missing values for each column

[49]:

CustomerID	Recency	Frequency	Monetary	R	F	M	RFM	RFM_Score	Recency_Log	Frequency_Log	Monetary_Log
12346.0	326	1	77183.60	1	1	4	114	6	5.786897	0.000000	11.253942
12347.0	2	182	4310.00	4	4	4	444	12	0.693147	5.204007	8.368693
12348.0	75	31	1797.24	2	2	4	224	8	4.317488	3.433987	7.494007
12349.0	19	73	1757.55	3	3	4	334	10	2.944439	4.290459	7.471676
12350.0	310	17	334.40	1	1	2	112	4	5.736572	2.833213	5.812338

Figure 5 Displaying RFM and Logs

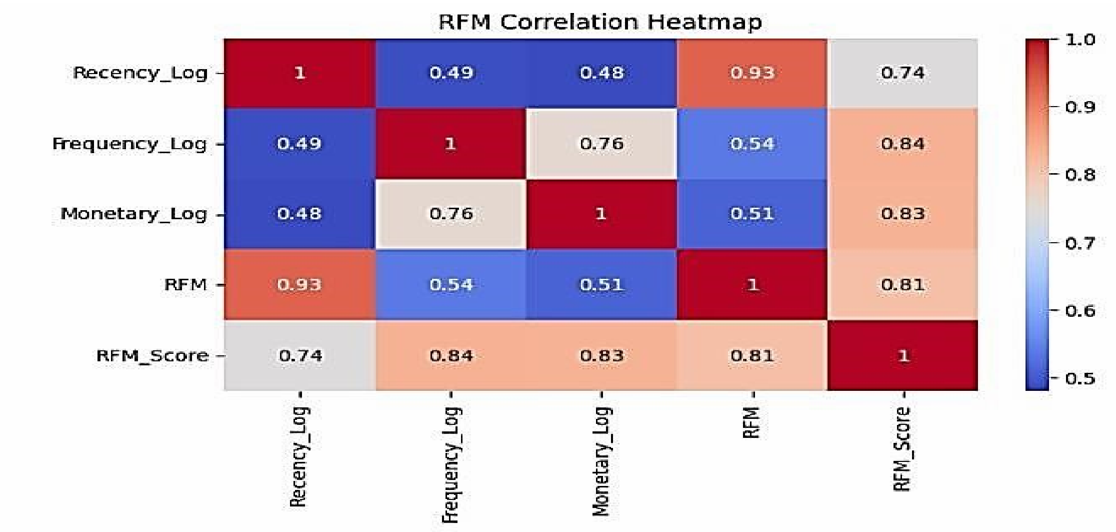

Figure 6 Visualizing Correlation Among Features

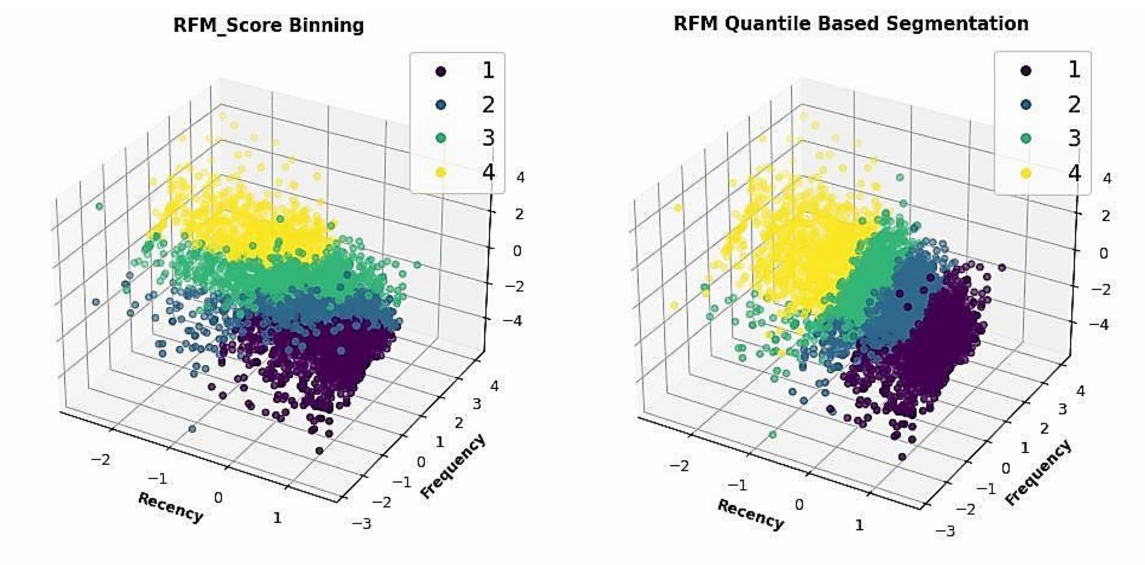

Figure 7 Displaying Binding Score Binning and RFM Quantile Based Segmentation

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 04 April 2025, pp: 34-43 www.ijaem.net ISSN: 2395-5252

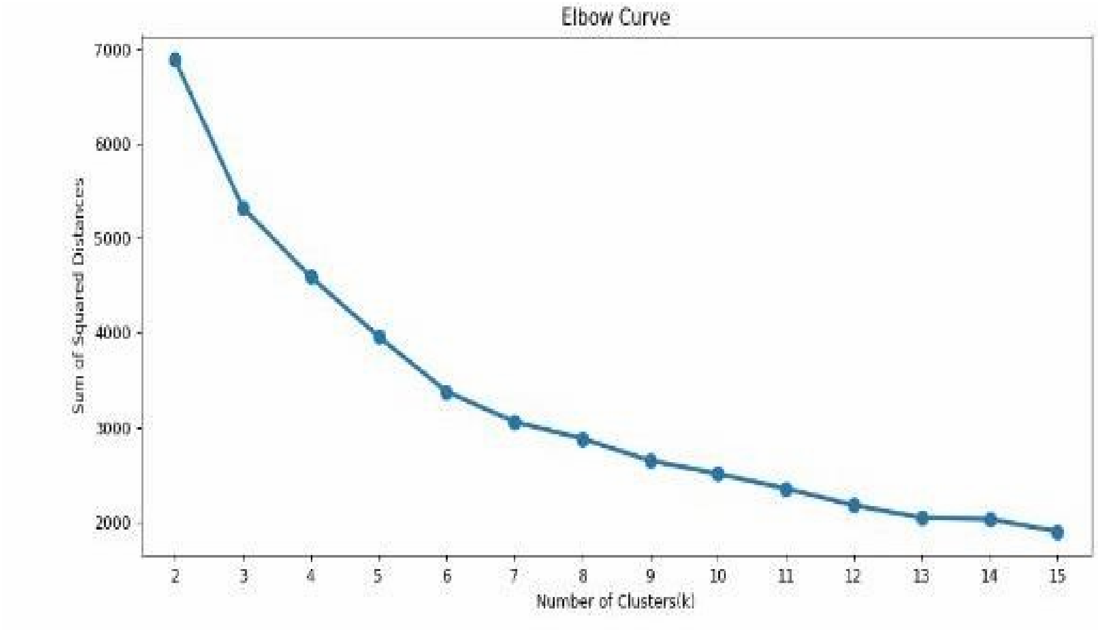

Figure 8 Plotting Elbow curve graph to find optimal K

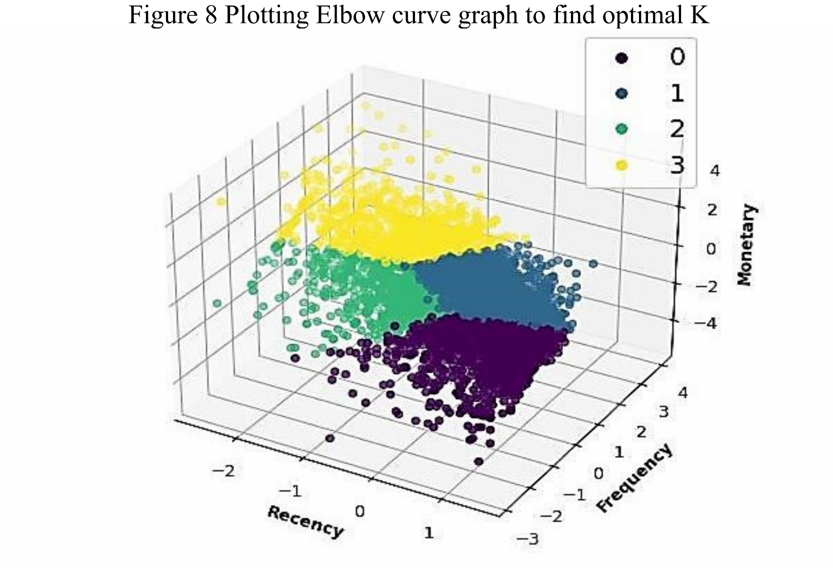

Figure 9 Displaying K-means based on Both Elbow curve and Silhouette score

Clusterer	Binning	Quantile Cut	K-Means	K-Means	K-Means	Agglomerative	Agglomerative	DBSCAN
Criterion	RFM Score Binning	RFM Quantile Cut	Elbow Curve	Silhouette Score	Elbow Curve & Silhouette Score	Dendogram (y=70)	Dendogram (y=50)	eps=0.2, min_samples=25
Segments	4	4	5	2	4	2	3	4

Figure 9.1 Summarizing the results obtained by clustering using various methods

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 04 April 2025, pp: 34-43 www.ijaem.net ISSN: 2395-5252

Binning	Last_visited	Purchase_frequency	Money_spent
1	93 to 274 days ago	Bought 7 to 20 times	Spent around 142 to 335 Sterling
2	31 to 114 days ago	Bought 19 to 41 times	Spent around 327 to 725 Sterling
3	16 to 65 days ago	Bought 46 to 98 times	Spent around 717 to 1613 Sterling
4	4 to 19 days ago	Bought 123 to 305 times	Spent around 2093 to 5398 Sterling

----

QuantileCut	Last_visited	Purchase_frequency	Money_spent
1	166 to 286 days ago	Bought 8 to 30 times	Spent around 156 to 486 Sterling
2	59 to 96 days ago	Bought 18 to 69 times	Spent around 355 to 1301 Sterling
3	23 to 39 days ago	Bought 28 to 118 times	Spent around 439 to 1887 Sterling
4	4 to 12 days ago	Bought 50 to 214 times	Spent around 822 to 3849 Sterling

----

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 04 April 2025, pp: 34-43 www.ijaem.net ISSN: 2395-5252

K-Means|2Cluster	Last_visited	Purchase_frequency	Money_spent
0	51 to 227 days ago	Bought 10 to 33 times	Spent around 187 to 562 Sterling
1	8 to 39 days ago	Bought 66 to 190 times	Spent around 1058 to 3342 Sterling

K-Means|4Cluster	Last_visited	Purchase_frequency	Money_spent
0	81 to 268 days ago	Bought 7 to 21 times	Spent around 144 to 367 Sterling
1	43 to 119 days ago	Bought 42 to 103 times	Spent around 705 to 1704 Sterling
2	4 to 17 days ago	Bought 120 to 308 times	Spent around 2072 to 5601 Sterling
3	9 to 27 days ago	Bought 20 to 52 times	Spent around 294 to 745 Sterling

hierarchical|2Cluster	Last_visited	Purchase_frequency	Money_spent
0	36 to 200 days ago	Bought 12 to 46 times	Spent around 219 to 739 Sterling
1	4 to 33 days ago	Bought 87 to 233 times	Spent around 1546 to 4020 Sterling

hierarchical|3Cluster	Last_visited	Purchase_frequency	Money_spent
0	4 to 33 days ago	Bought 87 to 233 times	Spent around 1546 to 4020 Sterling
1	30 to 163 days ago	Bought 29 to 63 times	Spent around 463 to 977 Sterling
2	49 to 243 days ago	Bought 6 to 18 times	Spent around 139 to 327 Sterling

hierarchical|2Cluster	Last_visited	Purchase_frequency	Money_spent
0	36 to 200 days ago	Bought 12 to 46 times	Spent around 219 to 739 Sterling
1	4 to 33 days ago	Bought 87 to 233 times	Spent around 1546 to 4020 Sterling

hierarchical|3Cluster	Last_visited	Purchase_frequency	Money_spent
0	4 to 33 days ago	Bought 87 to 233 times	Spent around 1546 to 4020 Sterling
1	30 to 163 days ago	Bought 29 to 63 times	Spent around 463 to 977 Sterling
2	49 to 243 days ago	Bought 6 to 18 times	Spent around 139 to 327 Sterling

Figure 9.2 Display of all result obtained from models used

IV.

V. RESULTS AND DISCUSSION

The study aims at clustering, segmentation customer using K-means clustering model, hierarchical clustering model, Density-based Spatial Clustering of Applications with Noise (DBSCAN) model and customer segmentation frameworks (RFM),This achievement indicate that the model can segment customers based on their behaviors and preferences, specific needs, product information, customer information fulfilling the primary aim of the research.

Also the study emphasizes the critical role of statistical analysis in achieving business objective, particularly in increasing profits through data driven insight. This focus on statistics is essential for understanding consumer interest and behavior, which are vital for effective marking strategies. The research evaluates various unsupervised machine learning clustering models, including K-means, hierarchical clustering, and DBSCAN. Each model was assessed for the insights it provided regarding customer segmentation, showcasing a comparative analysis of their effectiveness in real-world applications. The inclusion of a traditional model based on recently, frequency, and monetary (RFM) clustering in the analysis highlights the practical approach taken in evaluating the unsupervised models. This integration allows for a more comprehensive understanding of customer behavior and enhances the overall analysis

VI. CONCLUSION & RECOMMENDATION

The paper highlight effective customer segmentation strategies using clustering algorithms, it emphasizes the importance of personalized marketing for customer engagement. The analysis reveals insight from customer data to enhance organizational performance. Future enhancement in customer segmentation framework are suggested challenges were faced but improvement was identified for future development

The following suggestions should be taken into account in light of this work in the future. The software should be able to handle the following limitation encountered

i. Customer Life time analysis by using machine learning techniques like regression analysis or decision tress to predict the lifetime value of the customer. This can help you to identify customers that are likely to be the most profitable over time

ii. Customer Churn Prediction using machine learning algorithms like logic regression or decision tree to predict which customers are at risk to churning.

REFERENCE

[1]. A, A., Preemi, G., Sriram, S. R., Deva Shahila, D. F., Merlin Suji, M. A., & Sudeep Reddy, D. V. S. (2024). Smart Farming Based Early Classification of Paddy Blast Disease Using Adaptive Deep Learning Algorithm. 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT), 1, 49–53. https://doi.org/10.1109/ICCPCT61902.2024.10673091

[2]. Gupta, S., & Israni, D. (2024). Machine Learning based Customer Behavior Analysis and Segmentation for Personalized Recommendations. 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), 654–660. https://doi.org/10.1109/ICSSAS64001.2024.10760319

[3]. He, H., & Tan, Y. (2012). A two-stage genetic algorithm for automatic clustering. Neurocomputing, 81, 49–59. https://doi.org/10.1016/j.neucom.2011.11.001

[4]. Hicham, N., & Karim, S. (2022). Analysis of Unsupervised Machine Learning Techniques for an Efficient Customer Segmentation using Clustering Ensemble and Spectral Clustering. International Journal of Advanced Computer Science and Applications, 13(10). https://doi.org/10.14569/IJACSA.2022.0131016

[5]. Huang, H. (2024). Research on Customer Segmentation and Marketing Using Rubin

[6]. Index Based K Means Clustering. 1–4.

[7].

https://doi.org/10.1109/icdsns62112.2024.10690864

[8]. Loganathan, V., & Reddy, S. S. (2024). Customer Segmentation Using Machine Learning (pp. 204–209). https://doi.org/10.1201/9781003543633-32

[9]. Namvar, A., Ghazanfari, M., & Naderpour, M. (2017). A customer segmentation framework for targeted marketing in telecommunication. 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), 1–6. https://doi.org/10.1109/ISKE.2017.8258803

[10]. Nugroho, B. I., Rafhina, A., Ananda, P. S., & Gunawan, G. (2024). Customer segmentation

[11]. in sales transaction data using k-means clustering algorithm. Journal of Intelligent

[12]. Decision Support System, 7(2), 130–136. https://doi.org/10.35335/idss.v7i2.236

[13]. Rahmet, S., & Gürkan, F. (2024). Integration of Pre-processing and Machine Learning Methods for Improved Customer Segmentation. 2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP), 1–6. https://doi.org/10.1109/IDAP64064.2024.10710767

[14]. Shirole, R., Salokhe, L., & Jadhav, S. (2021). Customer Segmentation using RFM Model and K-Means Clustering. International Journal of Scientific Research in Science and Technology, 591–597. https://doi.org/10.32628/IJSRST2183118

[15]. Simbolon, I. N., & Friskila, P. D. (2024). analisis dan evaluasi algoritma dbscan (density-based spatial clustering of applications with noise) pada tuberkulosis. Jurnal Informatika Dan Teknik Elektro Terapan, 12(3S1), Article 3S1. https://doi.org/10.23960/jitet.v12i3S1.5206

[16]. Shen, M., Feng, B., Cheng, A., & Bi, Y. (2024). Customer Segmentation and Need Analysis

[17]. Based on Sentiment Network of Online Reviewers and Graph Embedding. Journal

[18]. of Mechanical Design 1–52. https://doi.org/10.1115/1.4067226

[19]. Ullah, A., Mohmand, M. I., Hussain, H., Johar, S., Khan, I., Ahmad, S., Mahmoud, H. A., & Huda, S. (2023). Customer Analysis Using Machine Learning-Based Classification Algorithms for Effective Segmentation Using Recency, Frequency, Monetary, and Time. Sensors, 23(6), Article 6. https://doi.org/10.3390/s23063180

[20]. Vardhan, V. S. S. H., & Sharma, A. (2023, December 1). Investigating the Degradation Behavior of Grain Refined WE43 Magnesium Alloy Produced by Friction Stir Processing for Medical Implant Applications. | EBSCOhost. https://doi.org/10.18311/jmmf/2023/41614

[21]. Wong, M. O., Ying, H., Yin, M., Yi, X., Xiao, L., Duan, W., He, C., & Tang, L. (2024). Semantic 3D reconstruction-oriented image dataset for building component segmentation. Automation in Construction, 165, 105558. https://doi.org/10.1016/j.autcon.2024.105558