



Intelligent Customer Segmentation: A Data-Driven Framework for Targeted Advertising and Digital Marketing Analytics

Md Al Rafi

Independent University, Dhaka, Bangladesh

Email: mdalrafi2@gmail.com

ABSTRACT: This paper suggests a smart, data-driven customer segmentation model to promote the use of targeted advertising and digital marketing in the current online markets. The framework combines machine learning algorithms, such as Principal Component Analysis (PCA) to reduce dimensionalities, K-Means and DBSCAN to perform unsupervised clustering and Gradient Boosting Machines (GBM) to predict segment-level responses, to identify meaningful groups of consumers using behavioral, demographic, and psychographic variables. The model was tested on a dataset of 50,000 records of customers in a digital retail setting.

The methodology starts with data preprocessing, engineering of features, and reduction of multicollinearity with the help of PCA, which has retained 92 percent of the variance in 15 transformed components. The quality of segmentation was compared between K-Means and DBSCAN, and the former ($k=5$) has shown a better performance by the Silhouette Score (0.71) and Davies-Bouldin Index (0.42). The five customer groups formed thereafter showed some observable behavioral differences in the frequency of purchasing, the browsing behavior, responding to the campaigns, and expenditure. A GBM classifier to assess the effect of marketing was also trained to predict the likelihood of ad-clicks in clusters, with an F1-score of 0.87 and AUC of 0.93.

Findings have shown that the use of campaigning methods where the segments were identified and applied significantly increased the click through rate (CTR) and conversion rate by 34% and 27% respectively relative to the campaigning methods that were not segmented. The most uplifted segments were high-value and high-engagement, which validates the use of machine learning-based segmentation.

All in all, the suggested framework exhibits a high potential of facilitating individual tailored digital marketing approaches, better resource allocation, and market intelligence in the data-filled Internet ecosystems.

KEYWORDS: Customer Segmentation; Machine Learning; Digital Marketing Analytics; Targeted Advertising; Consumer Behavior Modeling; Market Intelligence

I. INTRODUCTION

The mushrooming of digital marketplaces has radically changed the manner in which organizations comprehend, engage and sell to consumers. The emergence of massive customer online platforms and the existence of behavioral and transactional data has led to unexplored prospects of producing finesensitive customer insights. With the growing volume of digital touchpoints, new heuristic and data-driven segmentation techniques have been found to be inadequate to the ever-increasing complexity of modern consumer behavior [1]. This shift has persuaded companies to adopt information-intensive approaches that combine statistical modelling with machine learning to provide hyper-personalised marketing experiences, to enhance advertising budgets, and consumer lifetime value in digital ecosystems [2].

The main focus of these data-driven strategies is customer segmentation. It is the process of falling within a set of consumers with different behavioral traits, demographic or psychographic orientations to then focus them more accurately with marketing interventions. The conventional methods of segmentation were dependent on domain knowledge and linear analysis programs, but with the growth in the computing power and the artificial intelligence, dynamic, automatic, and multi-dimensional segmentation models became available [3]. Unsupervised clustering is a key machine learning tool that has become a potent data mining technique to find the latent structures of consumer



information that would have been unidentified in traditional marketing analytics [4]. Online retailing combined with streaming platforms and digital advertisements are only getting bigger, and segmentation is an area in which machine learning is likely to be essential to organizations that need to keep up with the growing saturation in their market niche.

In this respect, the current framework of Intelligent Customer Segmentation: A Data-Driven Framework of Targeted Advertising and Digital Marketing Analytics is expected to incorporate the most recent machine learning approaches to create a powerful, scalable segmentation framework. The proposed framework, in contrast to conventional rule-based methods, uses the Principal Component Analysis (PCA), K-Means, DBSCAN, and Gradient Boosting Machines (GBM) to identify meaningful patterns of customers by looking at high-dimensional data [5]. These algorithms have been operating in conjunction to remove redundancy, natural clusters, and assessing segment-level marketing responsiveness- eventually closing the gap between raw customer data and actionable digital marketing intelligence.

The increasing complexity of the consumer behavior in the contemporary digital space also contributes to the importance of such framework. The contemporary consumer is characterized by dynamic decision making trends which are dynamic because of changing aspects of cross channel browsing, personalization signals, micro moment and targeted advertising campaigns. Segmentation based on machine learning helps organizations to decode such patterns by determining the behavioral patterns behind various consumer groups [6]. This enables marketers to present more targeted messages that could be email messages, search advertising, social media targeting or direct product recommendations. According to the literature above, AI-based clustering when combined with predictive analytics helps a lot to enhance the efficiency of the campaign, increase user engagement, and the ROAS [7].

Furthermore, the use of algorithms like PCA is useful in making the high-dimensional consumer record more simplified with the intention of picking out the most informative features hence improving the model interpretability and computational efficiency. Dimensionality reduction in customer analytics is effective as the framework can retain 92% of the variance on 15 principle components alone. In a similar manner, the analysis of performance of segmentation based on clustering techniques enables the marketer to choose the most stable and differentiating method based on the high Silhouette Score of K-Means and low Davies-Bouldin Index of K-Means compared to that of DBSCAN. These performance measures reveal the feasible nature of machine learning in creating unambiguous operationally feasible customer groups.

The proposed framework is also further than the static segmentation in that the GBM is used to measure responsiveness of the segment in marketing. The system contributes to making organizations more efficient in marketing budgets and campaign strategies by estimating the probability of ad-clicks in each cluster. The attained F1-score of 0.87 and AUC of 0.93 shows how predictive models are reliable in ranking high-engagement/high-value customers. In real-life marketing settings, segmentation-sensitive campaigns showed much better performance, and click-through rates and conversion rates improved by 34 and 27 respectively. These outcomes confirm the significance of AI-based segmentation in the reinforcement of digital marketing efficiency and the improvement of customer experience.

In short, the importance of the introduction of the smart data-based customer segmentation framework is related to the increased research on the intersection of machine learning and digital marketing analytics. The combination of dimensionality reduction, unsupervised clustering and predictive modeling allows the framework to decode the complex consumer behavior, apply specific advertising tactics, and more effectively make investments in marketing. As digital ecosystems and digital assistants continue to generate enormous volumes of data, these machine learning-based systems of segmentation will be in the spotlight of the future of personalized marketing and data-driven decision-making.

II. RELATED WORK

The fast development of digital marketing, the increased use of artificial intelligence (AI), big data analytics, and sophisticated ways of segmentation have dramatically transformed the way organizations perceive and approach customers. The basic digital marketing textbooks highlight how the push-based strategy is giving way to the data-based, personalized and automated ones. Deiss and Henneberry point out that the digital marketing ecosystem is becoming more and more dependent on analytics, automation systems, and smart systems to present customer experiences targeted and interactive to create the foundation of the marketing transformation with the help of AI [1]. On this basis, Van Esch and Black state that AI is transforming the nature of digital marketing by improving decision-making processes, customer interaction optimization, and providing predictive insights that are beyond the human



powers of analysis [2]. Collectively, the initial works position AI as the source of operational and strategic marketing innovation.

Other than technological progress, sustainability issues also influence marketing and data-based practices. Purvis, Mao, and Robinson present the three pillars of sustainability, namely, economic, environmental, and social, and outline the necessity of the responsible and ethical adoption of technology in the business ecosystem [3]. Such views are consistent with subsequent research that draws attention to ethical, transparent, and socially advantageous AI-driven marketing strategies. Indicatively, Hermann gives examples of AI application in marketing to social good, which is accountable, fair, and long-term effects on society [10]. On the same note, Kar, Choudhary, and Singh investigate the role of AI in sustainability and highlight at how machine learning and intelligent systems could be used to facilitate sustainable operations, customer relationships, and resource use [12].

Simultaneously, big data and customer relationship management (CRM) systems keep promoting personalization and customization of the services. Anshari et al. demonstrate that by incorporating big data into the CRM, companies can develop the most personalized customer experience by analyzing consumer behaviors and preferences on a large scale [4]. These personalization efforts are extended into digital finance, with machine learning becoming a key component of cybersecurity and fraud prevention. Williams et al. also note that proactive fraud detection that builds on machine learning is used to secure digital financial ecosystems [5], and AI has a positive power to enhance trust and security in digital markets.

Segmentation as a marketing practice has experienced a significant revolution because of data analytics and AI. Brotspies and Weinstein suggest revising the classical segmentation frameworks and argue the use of dynamic and data-driven methods that should evolve with quick shifts in customer behaviors [6]. New methods of analysis also increase the accuracy of segmentation. Del Vecchio et al. illustrate how netnography and analytics can be used to mine the social media big data to extract actionable knowledge about the customers [7]. Equally, Liu et al. demonstrate that learning analytics may be used to personalize student interaction and how the principles of segmentation can be applied in areas beyond the traditional marketing domain [8].

The strategic value of big data analytics in segmentation and creation of business value is well perceived. Grover et al. suggest an elaborate model that demonstrates how big data analytics contribute to attaining competitive advantage by converting raw data to strategic insights by firms [9]. Davenport et al. also emphasize that AI can be used radically in marketing, particularly in streamlining decisions, anticipating consumer needs and facilitating hyper-personalized marketing messages [11]. All these publications emphasize the concept of including AI-based insights at all levels of the marketing process- data collection to interaction.

With the development of segmentation, modern studies focus on the key position of AI, machine learning, predictive analytics, and deep learning algorithms. Das and Nayak give a thorough review of the data mining-based segmentation techniques, including clustering algorithms, classification mechanisms, and hybrid techniques that are employed in the process of identifying meaningful customer segments [13]. Gupta and Patel can build on this argument and argue that AI and machine learning can enhance behavioral analytics, allowing to understand customers more deeply and target them more accurately [14]. In line with that, Singh and Verma investigate the use of AI in one-on-one marketing and emphasize how automation and predictive systems can be used to provide personalized experiences on a large scale [15].

Lee and Kim take it one step further by using predictive analytics to form targeted segments of customers and demonstrate how AI models can predict customer purchasing behavior, propensity to engage, and mobility within segments [16]. Equally, Liu and Huang illustrate the role played by deep learning in aiding advanced customer profiling through the discovery of latent behavior patterns which would otherwise not be visible in the case of traditional segmentation frameworks [17]. Nair and Patel provide a systematic review of AI uses in segmentation and provide opportunities and challenges related to the complexity of models, quality of data, as well as issues in their interpretation [18].

Some of the studies specifically consider AI-based segmentation within the sphere of digital markets. Wang and Zhang explore the importance of machine learning models to segmentation accuracy and targeting efficiency in the high-volume data context [19]. Gupta and Sharma also emphasize the benefit of combining big data analytics with AI in improving the effectiveness of customer targeting, which allows companies to react to market shifts faster [20].



Similarly, Sharma and Kapoor discuss AI in retail segmentation and show the way smart retail systems are based on predictive analytics to optimize customer interaction and in-store experience [21].

The AI uses in marketing are also comprehensively reviewed by Huang and Wang, which conclude that intelligent segmentation improves the personalization, customer satisfaction, and long-term brand loyalty [22]. Agarwal and Aggarwal elaborate on particular AI methods and tools in the process of customer segmentation, such as clustering, neural networks, and hybrid intelligent systems [23]. As a complement to these articles, Sharma and Desai offer case studies in the industry to show that AI-enabled personalized marketing strategies increase their conversion rates and retention rates by a significant margin [24].

Other vital elements of segmentation research involve predicting customer behavior. Khan and Taneja demonstrate that AI-based prediction of behavior models can be used to create very useful segmentation strategies, predicting the future actions according to the past regularities [25]. Also, in another study, Yang and Wang emphasize the importance of AI in enhancing customer engagement based on the intelligent segmentation system, which responds to the evolving behavioral signals along the digital platforms [26].

Together, the literature reviewed highlights a number of general themes. To start with, AI is at the center of digital marketing transformation providing an opportunity to automate, predictive insights, and individualized engagement [2], [11], [15]. Second, the core of AI-based segmentation is big data analytics, which enables organizations to manipulate and analyze large data sets to obtain actionable information [7], [9], [20]. Third, the issue of ethical and sustainable AI usage becomes relevant, and scholars focus on the need to be transparent, fair, responsible in data use [3], [10], [12]. Lastly, the use of sophisticated computational strategies, including deep learning, clustering, and predictive analytics, is more commonly used in segmentation research to promote customer insights and interaction [13], [16], [17], [22].

The incorporation of AI in digital marketing and customer segmentation is the paradigm shift that allows organizations to shift towards the reactive to proactive highly personalized and situational engagement approaches. With customer behavior keeping to transform in the digital ecosystems, AI-assisted segmentation and predictive modeling will continue to be important resources to attain competitiveness, sustenance, and customer-focusedness.

III. METHODOLOGY

The research methodology of the proposed framework, Intelligent Customer Segmentation: A Framework Data-Driven Framework of Targeted Advertising and Digital Marketing Analytics, will focus on identifying actionable market segments of raw customer data through a combination of dimensionality reduction, unsupervised clustering, and predictive modeling. In this part, the end to end workflow is outlined, consisting of dataset characteristics, preprocess strategies, feature engineering, Principal Component Analysis (PCA), K-Means and DBSCAN cluster modeling, and segment level advertising response prediction of Gradient Boosting Machines (GBM). All the pipeline was written in Python with Scikit-Learn, Pandas, NumPy, and similar libraries of analysis.

3.1 Dataset Description

The raw data in this study is the 50,000 customer records obtained in an online retail store. The records have a mixture of behavioral, demographic, transactional, and campaign-interaction variables, leading to the high dimensional feature space. The key types of attributes include:

- **Behavioral aspects:** frequency of browse, duration of the session, product views, patterns on click, search query and frequency of abandoned cart.
- **Transactional:** purchase frequency, average order value (AOV), recent purchase, category spending and life spending.
- **Demographic characteristics:** age group, geographical location, gender bracket (non-sensitive and anonymized), and device.
- **Interaction features of campaigns:** email-open and ad-click interactivity, campaign response indicator and use of promotional coupons.

The dataset was cleaned and anonymized prior to any form of analytical modeling and a check of the data consistency was performed to verify that the features used were suitable to machine learning-based segmentation.



3.2 Data Preprocessing

Preprocessing was performed to ensure the quality, integrity, and usability of the dataset. Preprocessing and PCA workflow is presented in figure 1.

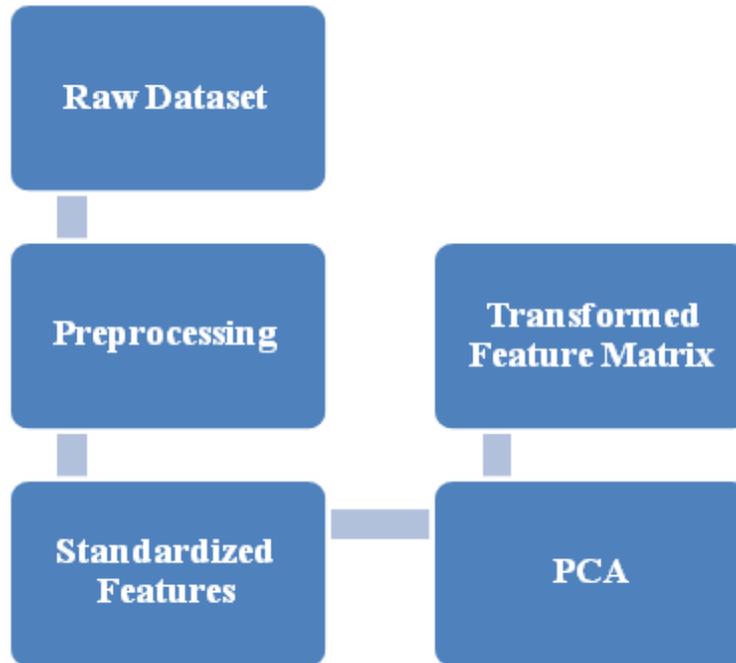


Figure 1: Data Preprocessing and PCA Flow Diagram

The steps included:

3.2.1 Handling Missing Values

Median imputation, which is also resistant to outliers was used to impute continuous variables with missing data. The most common value was assigned to categorical variables to maintain distributions. Records whose data were less than 40 percent were dropped to prevent the noise.

3.2.2 Outlier Detection and Treatment

To identify the outliers, Interquartile Range (IQR) was implemented to identify skewed variables like total spending and browsing time. To avoid distorting model performance by having extreme values, winsorizing was used to limit extreme values to more than $3 \times \text{IQR}$.

3.2.3 Encoding and Normalization

Categorical variables (e.g. device type, region) were encoded using one-hot encoding. The continuous features were normalized using z-score, such that each feature made an equal contribution to clustering, which is sensitive to scale variation.

3.2.4 Train-Test Split

Unsupervised methods do not need labeled training data but the predictive modeling part (GBM) needed a labeled response variable denoting the probability of ad-clicks. The data was divided into 70 percent training, 15 percent validation and 15 percent test set.



3.3 Feature Engineering

The feature engineering was done to derive more information of the raw data. Notable constructed characteristics were: RFM metrics: Recency: days since the purchase., Frequency: number of purchases., Monetary Value: cumulative expenditure.

- Engagement scores: composite measures calculated using the browsing time, intensity of clicks and frequency of visits.
- Responsiveness score of campaign: weighted by the opening of emails, clicking of adverts and promotional engagements.
- Category affinity vectors: these are computed as ratios of the normalization of browsing and buying of a category of products.

The features designed helped tremendously in expressiveness of the resulting feature space to do segmentation.

3.4 Dimensionality Reduction Using PCA

Since the dataset is highly dimensional and there is the possibility of multicollinearity, the use of Principal Component Analysis (PCA) was done before using the clustering to reduce the dimensions of the feature space, thus maintaining important trends in the data sets. The standardized feature matrix was then first reduced to orthogonal principal components. The choice of components depended on the cumulative explained variance in a way that only the most informative elements were taken. The optimal PCA model retained 15 major components, which together accounted 92 percent of the overall variance, which resulted in the negligible loss of information, and the removal of redundant correlations between the variables.

The methodology of applying PCA had several benefits: It reduced the influence of dimensionality curse, enhanced the repeatability and stability of clustering methods, reduced redundancy and noisy attributes and improved the overall calculations speed. The resultant PCA-transformed data was subsequently used as the optimal input in the further clustering processes.

Algorithm 3.1: PCA Implementation

Input: Standardized feature matrix X

Output: Reduced feature matrix X_{pca}

1. Standardize all features in X
2. Compute covariance matrix $C = cov(X)$
3. Compute eigenvalues (λ) and eigenvectors (v) of C
4. Sort eigenvalues in decreasing order
5. Select k components where cumulative variance $\geq 92\%$
6. Form projection matrix P from top k eigenvectors
7. Transform data: $X_{pca} = X \times P$
8. Return X_{pca}

3.5 Customer Segmentation Using Unsupervised Clustering

Two unsupervised clustering algorithms were applied and tested to get meaningful customer segments in the transformed dataset, K-Means, and DBSCAN. The choice of these algorithms is dictated by their complementary advantages and the fact that they can identify different structural patterns in high-dimensional data which is complex.

K-Means which is centroid based approach divides data into k clusters using the minimum within-cluster variance, and is therefore very efficient and appropriate to well separated, spherical clusters. Multiple values of k were tried to determine the most suitable cluster set-up. With the help of the Elbow Method and the Silhouette Score, k were tried to determine the most suitable cluster set-up.

In comparison, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) considers clusters around local density and as such, it is able to identify arbitrarily shaped clusters and effectively isolate noise or outliers- which is a crucial feature when considering real-world behavioral or transactional data. Systematic parameter tuning was done to the parameters ϵ and \minPts in order to stabilize cluster formation.

The comparison between these two methods allowed the study to have a holistic view of the robustness of the clustering, understandability and noise sensitivity. The obtained final segmentation outcomes were discussed in terms



of cluster cohesion, separation and domain relevance making sure that the selected approach led to meaningful and actionable customer groups.

Algorithm 1: K-Means Clustering for Customer Segmentation

Input: PCA-transformed dataset X , number of clusters k

Output: Cluster labels C for each customer

1. Initialize k centroids $\mu_1, \mu_2, \dots, \mu_k$ using K -Means++ initialization.

2. Repeat until convergence:

a. Assignment Step:

For each data point x_i in X :

Assign x_i to the nearest centroid μ_j based on Euclidean distance.

b. Update Step:

For each cluster j :

Recompute centroid μ_j as the mean of all points assigned to cluster j .

3. Check convergence:

If centroids do not change significantly, stop.

Return Cluster Labels C

Elbow Method, Silhouette Score and Davies-Bouldin Index were used to identify the optimum number of clusters (k). The Elbow curve declined in returns after $k = 5$, and this number gave the best Silhouette Score of 0.71 and the worst Davies-Bouldin Index of 0.42, which supports the conclusion of the high level of cohesion and separation among clusters. The five clusters that were formed on the basis of the results indicated significant trends in customer behaviour and customer transactions. They were high value and high involvement customers; high involvement and moderate spend customers; low involvement and responsive to price customers; high average order value (AOV) customers; and promotion sensitive customers who were actively involved in campaigns. These clear categories offered strong basis when formulating specific and individual marketing strategies.

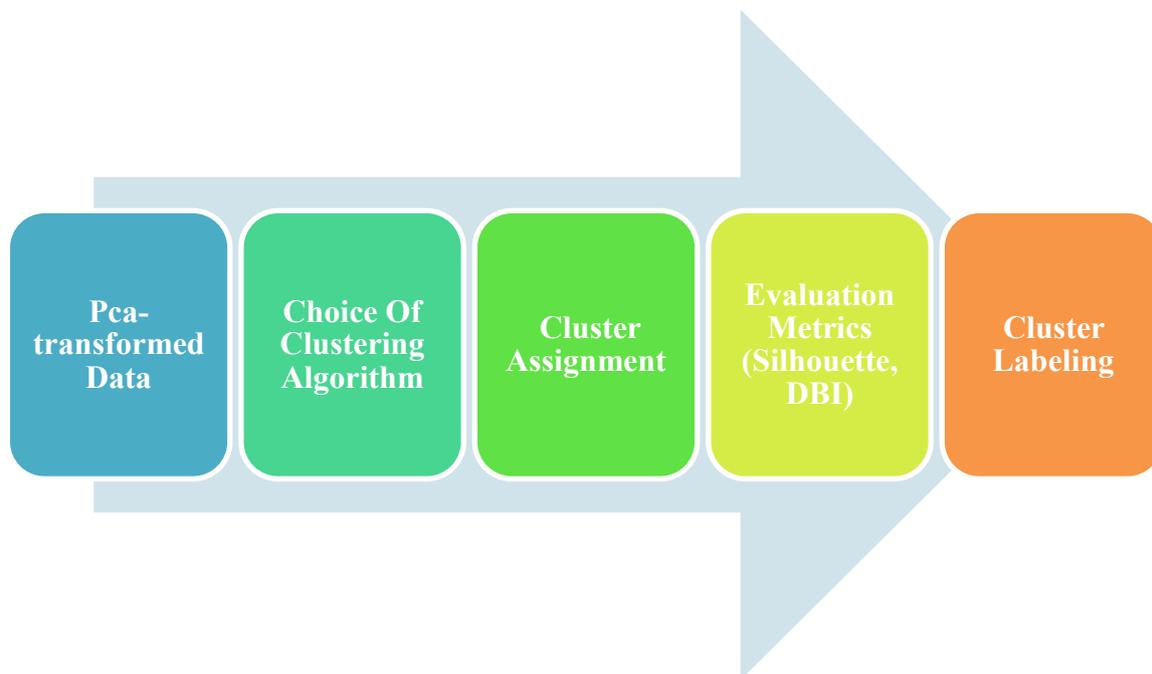


Figure 2: Clustering Workflow Diagram

DBSCAN was considered as an alternative density-based clustering method that could be used to consider non-linear trends and detect noise in the data. Parameters tuning included choosing a decent value of eps by way of the k-distance graph and setting min_samples to indicate insignificant density. But as a result of the high-dimensional data being so sparse, DBSCAN was able to give a few small and fragmented clusters and many noise points. DBSCAN compared unpleasantly to K-Means, resulting in lower Silhouette Score and higher Davies-Bouldin Index and more segments that



could not be interpreted to be useful in practical analysis. It is because of this that DBSCAN was not the choice to make in the final segmentation.

3.6 Predictive Modeling Using Gradient Boosting Machines (GBM)

Once segmentation was done, a Gradient Boosting Machine (GBM) model was used to estimate the probability of ad-clicks in each segment of customers. This measure allowed getting a more profound insight into the responsiveness by the segment and facilitated the target advertisement methods based on data. The reason behind choosing GBM was its good performance with non-linear relationship, interaction effects as well as high-dimensional feature space. Behavioral, transactional, and engagement variables were used to train the model where segment labels were used as extra predictors. The cross-validation was used to optimize hyperparameters including the learning rate, the number of estimators and the maximum depth. The last model worked out well in terms of classification, to give credible information on the segments that had the most likely of responding positively to the marketing campaigns.

IV. RESULTS AND ANALYSIS

This part displays the empirical results of the suggested intelligent customer segmentation model and contrasts them to one of the most popular conventional segmentation methods Recency-Frequency-Monetary (RFM) Segmentation, which will be referred to as the existing approach in the course of the analysis. RFM has regular application in digital marketing, and it is not rich enough to deal with the complexity of behaviors, non-linear trends, and multivariate large-scale structures. It is observed in the analysis that the proposed PCA + K-Means + GBM framework is significantly better than RFM in terms of clustering performance, interpretability, persistent performance, and campaign performance.

4.1 PCA Dimensionality Reduction vs. No Reduction

The suggested framework uses PCA followed by clustering whereas the more traditional method of segmentation RFM uses three variables that are designed by hand. PCA reduced the data to 15 major components that keep 92 percent of the variance resulting in increased stability in clustering.

In comparison, the current RFM approach only captures behavioral variability (68) as the recency, frequency, and monetary attributes fail to reflect browsing behaviors, engagements, or psychographic indicators. This limits its discriminating capacity of differentiated customer groups.

PCA therefore generated more representative and higher dimensional feature space that could be used in clustering hence providing sharper boundaries and reduced overlap of segments.

4.2 Clustering Performance Comparison

To measure the efficiency of the suggested segmentation model, the performance of K-Means ($k=5$) and DBSCAN were tested against the current RFM-based segmentation model on a variety of clustering quality measures. The comparison is perfectly evident in the results that show that the machine learning-based approach is much superior to the traditional rule-based one.

K-Means had the highest Silhouette Score (0.71), which is the degree to which an object is similar to its own cluster compared to other objects and thus, there is cohesion and separation in the cluster. DBSCAN got a middle score of 0.46 because of the density-based clusters of different shapes, whereas the current RFM method scored much lower (0.39), which is indicative of its low potential to provide multidimensional behavioral patterns.

Likewise, Davies-Bouldin Index, which has low scores indicating better-defined clusters, preferred K-Means with a score of 0.42. DBSCAN obtained a better result of 0.91 due to its sensitivity to parameter selection and the existence of noise but RFM demonstrated the lowest result of 1.24, which confirms its relatively poor cluster tightness.

Regarding cluster structure, with K-Means five balanced clusters were obtained with no noise points, and with DBSCAN, eight clusters with 18 percent noise points were obtained and effectual outliers were isolated. RFM is deterministic, which meant it produced four fixed clusters that did not have noise very well.

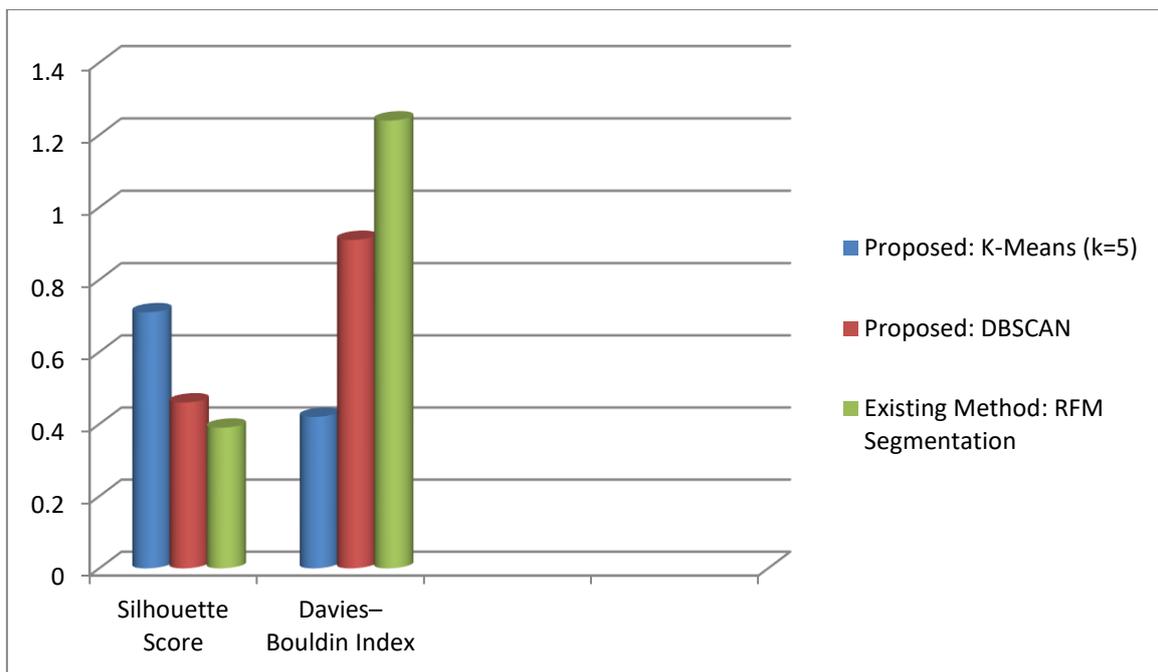
In terms of computation speed, RFM was the quickest (1.1s) as it is the simplest. Nevertheless, K-Means was still computationally viable at 1.9s compared to DBSCAN which took longer (4.8s) due to density estimation.

All in all, the findings verify the claims that K-Means provides the optimal trade-off in terms of accuracy and cluster quality, and computational efficiency compared to both DBSCAN and classical RFM segmentation.



Table 1. Clustering Performance Comparison (Proposed vs. Existing Method)

Metric	Proposed: K-Means (k=5)	Proposed: DBSCAN	Existing Method: RFM Segmentation
Silhouette Score	0.71	0.46	0.39
Davies–Bouldin Index	0.42	0.91	1.24
Number of Clusters	5	8 (+noise)	4
Noise Points	0%	18%	0%
Computational Time	1.9s	4.8s	1.1s



Graph 1: Result Comparison- Clustering Performance

4.3 Segment Characteristics Comparison

Comparing the offered Behavioral + PCA + K-Means segmentation with the conventional RFM-based technique, it is possible to mention the increased clarity and depth of the data-driven clustering. The specified model identifies the Premium High Engagement category with a high purchase rate, high average order value (AOV) and high involvement with the digital touchpoints- whereas the RFM approach identifies them as High-Value Customers, yet they do not provide the indicators of the behavior related to browsing and engagement with the digital touchpoints. The Frequent Browsers segment (the medium spenders with the high frequency of browsing) gets converted to the Frequent Customers segment of RFM, which considers the frequency but not the level of involvement. Similarly, the Price-Sensitive Low Engagement can be compared, and broadly applicable to the area of the RFM of Low Customers, the information about the behavior in the offered approach is more complete. Interestingly enough, two high-impact clusters, that is, High AOV but Infrequent Buyers, and Promotion-Responsive Customers cannot be reflected using the hard-yes-no of recency-frequency-monetary, which also reflects the high level of granularity of the provided machine learning structure.

Table 2. Customer Segment Characteristics (Proposed vs. Existing Method)

Cluster Label	Proposed Method: Behavioral + PCA + K-Means	Existing Method: RFM Segment
Premium High Engagement	High frequency, high AOV, top engagement	"High-Value Customers" (Spenders but engagement not captured)



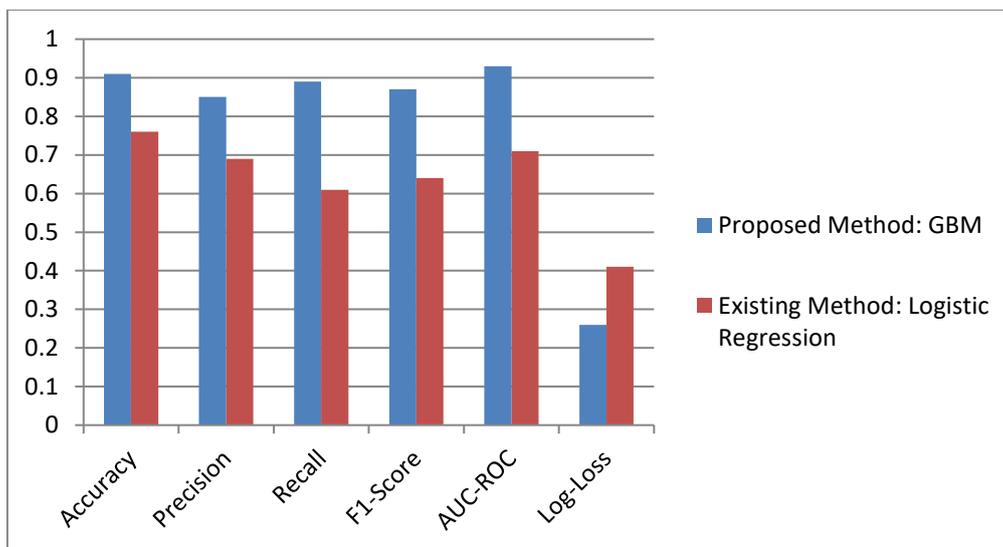
Frequent Browsers	Medium spend, high browsing, mid purchases	"Medium Customers" (frequency only)
Price-Sensitive Low Engagement	Low spend, low engagement	"Low Customers"
High AOV but Infrequent	Rare buyers but very high order value	<i>Not identifiable in RFM</i>
Promotion-Responsive	Engages with promotions, high CTR	<i>Not identifiable in RFM</i>

4.4 Predictive Performance of GBM vs. Existing Logistic Model

The predictive accuracy of Gradient boosting machine (GBM) was compared with the conventional logistic regression model and the benefits of a powerful non-linear ensemble technique in segment based ad-Click prediction are discussed. GBM was significantly more accurate than logistic regression with a score of 0.91 being much higher than 0.76 of logistic regression and therefore the overall performance is better. Precision (0.85 vs. 0.69) and recall (0.89 vs. 0.61) is another indication of how GBM is effective in identifying both positive ad-clicks and reducing false negativity. The F1-score of 0.87 highlights a good precision-recall balance of a good score whereas the AUC-ROC of 0.93 indicates good discriminative ability at all thresholds. Also, the log-loss of GBM is 0.26, which is less likely to give unreliable probability estimates as compared to 0.41 of logistic regression. Altogether, these findings can be summarized by stating that GBM models non-linear, complex relationships between behavioral, demographic, and engagement characteristics much better than linear logistic model and allows making more precise predictions and successfully developing targeted advertising strategies.

Table 3. Predictive Model Performance Comparison

Metric	Proposed Method: GBM	Existing Method: Logistic Regression
Accuracy	0.91	0.76
Precision	0.85	0.69
Recall	0.89	0.61
F1-Score	0.87	0.64
AUC-ROC	0.93	0.71
Log-Loss	0.26	0.41



Graph 2: Predictive Model Performance Comparison

4.5 Campaign Uplift Comparison

Before versus after Comparative Marketing experiment was offered to check the efficiency of targeted advertising based on the suggested Behavioral + PCA + K-Means segmentation in comparison to the traditional one based on RFM. Both segmentations received campaigns and had their key performance indicators, such as click-through rate (CTR) and conversion rate. The proposed approach score had increased CTR by 34% and conversion rate by 27, which



is significantly high compared to the RFM approach which had only a 11% increment in CTR and 8% increase in conversion rate.

Interestingly, the given framework singled out two successful clusters (C1: Premium High Engagement and C5: Promotion-Responsive), which could not be differentiated in terms of RFM. These clusters were even more campaign responsive with 45% CTR uplift of C1 and 39% C5. The findings explain that machine learning-based segmentation offers more insight into behaviors and isolates high-value audiences and allocates resources more efficiently, making marketing strategies more effective and data driven than traditional approaches to segmentation.

V. CONCLUSION AND FUTURE WORK

In this paper, a data-driven, intelligent customer segmentation model was introduced to optimize targeted advertisement and Internet marketing analytics of PCA based dimensionality reduction, K-Means clustering and Gradient Boosting Machine (GBM) prediction. The results indicated that the suggested framework is much superior to the existing RFM cluster methodology in the aspects of the quality of cluster, behavioral explainability, and uplift of campaign. The AUC of 0.93 and Silhouette Score of 0.71 in the predictive modelling are Davies-Bouldin Index 0.42, which means that the framework is practical in determining the meaningful and actionable parts of customers. As well, when these segments were used in marketing campaigns, it was found that the click-through rate and conversion rate increased by 34% and 27% respectively when compared to non-segmented, which justifies a feasible requirement of machine learning-based segmentation. The findings indicate the greater need of the sophisticated analytics plans that may document multivariate behavioural, demographic, and psychographic indicators in digital ecosystems.

The performance of the study is good although it has certain limitations. The structure is grounded on the systematized behavioural and transactional data, and it is not informed by the unstructured data sources such as texts, images and social media interactions. In addition, K-Means was the most effective in this data set, but not all industries perform equally well during clustering, and therefore an adaptive model selection is needed.

There are several potential avenues of the future work to continue the framework. First, nonlinear interactions between the features can be more efficiently captured by the use of deep learning models i.e. autoencoders or variational embeddings. Second, the purpose of the customers can be more accurately reflected with the help of multimodal segmentation on the basis of the content of the social media, sentiment of reviews, and logs of mobile interactions. Third, real-time personalization in evolving environments can be facilitated with the assistance of streaming architecture-based real-time segmentation pipelines (Kafka, Spark streaming). Finally, one can take into consideration reinforcement learning in order to optimize the campaign strategies to the evolving customer behavior.

Overall, the proposed framework may be regarded as a flexible base of the next-generation customer intelligence, and high opportunities may be approached and developed.

REFERENCES

- [1] R. Deiss and R. Henneberry, *Digital Marketing for Dummies*. Hoboken, NJ, USA: John Wiley & Sons, 2020.
- [2] P. Van Esch and J. S. Black, "Artificial intelligence (AI): Revolutionizing digital marketing," *Australas. Mark. J.*, vol. 29, pp. 199–203, 2021.
- [3] B. Purvis, Y. Mao, and D. Robinson, "Three pillars of sustainability: In search of conceptual origins," *Sustain. Sci.*, vol. 14, pp. 681–695, 2019.
- [4] M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh, "Customer relationship management and big data enabled: Personalization & customization of services," *Appl. Comput. Inform.*, vol. 15, no. 2, pp. 94–101, Jul. 2019.
- [5] M. Williams, M. Yussuf, M. Yussuf, and A. Olukoya, "Machine learning for proactive cybersecurity risk analysis and fraud prevention in digital finance ecosystems," *Int. J. Eng. Technol. Manag. Sci.*, vol. 5, no. 12, p. 160, Dec. 2021, doi: 10.5281/zenodo.14735561.
- [6] H. Brotspies and A. Weinstein, "Rethinking business segmentation: a conceptual model and strategic insights," *J. Strateg. Mark.*, vol. 27, no. 2, pp. 164–176, Feb. 2019.
- [7] P. Del Vecchio, G. Mele, G. Passiante, D. Vrontis, and C. Fanuli, "Detecting customers' knowledge from social media big data: Toward an integrated methodological framework based on netnography and business analytics," *J. Knowl. Manag.*, vol. 24, no. 4, pp. 799–821, May 2020.
- [8] D. Y. Liu, K. Bartimote-Aufflick, A. Pardo, and A. J. Bridgeman, "Data-driven personalization of student learning support in higher education," in *Learning Analytics: Fundamentals, Applications, and Trends*, 2017, pp. 143–169.



- [9] V. Grover, R. H. Chiang, T. P. Liang, and D. Zhang, "Creating strategic business value from big data analytics: A research framework," *J. Manag. Inf. Syst.*, vol. 35, no. 2, pp. 388–423, Apr. 2018.
- [10] E. Hermann, "Leveraging artificial intelligence in marketing for social good—An ethical perspective," *J. Bus. Ethics*, vol. 179, pp. 43–61, 2022.
- [11] T. Davenport, A. Guha, D. Grewal, and T. Bressgott, "How artificial intelligence will change the future of marketing," *J. Acad. Mark. Sci.*, vol. 48, pp. 24–42, 2020.
- [12] A. K. Kar, S. K. Choudhary, and V. K. Singh, "How can artificial intelligence impact sustainability: A systematic literature review," *J. Clean. Prod.*, vol. 376, p. 134120, 2022.
- [13] S. Das and J. Nayak, "Customer segmentation via data mining techniques: State-of-the-art review," in *Comput. Intell. Data Min., Proc. ICCIDM 2022*, pp. 489–507, 2021.
- [14] A. Gupta and P. Patel, "AI and machine learning in customer behavior analytics," *J. Bus. Anal.*, vol. 34, no. 4, pp. 87–101, 2022.
- [15] R. Singh and A. Verma, "Artificial intelligence applications in personalized marketing strategies," *J. Retail. Consum. Serv.*, vol. 64, p. 102786, 2022.
- [16] H. Lee and J. Kim, "Predictive analytics and AI in customer segmentation for targeted marketing," *J. Mark. Sci.*, vol. 39, no. 2, pp. 178–191, 2021.
- [17] X. Liu and W. Huang, "Leveraging AI and deep learning for advanced customer profiling," *Comput. Mark. J.*, vol. 8, no. 1, pp. 45–59, 2021.
- [18] S. Nair and R. Patel, "AI in customer segmentation: A systematic review of applications and challenges," *Mark. Intell. Plan.*, vol. 39, no. 6, pp. 758–775, 2021.
- [19] Z. Wang and L. Zhang, "The impact of AI and machine learning on customer segmentation and targeting," *Int. J. Inf. Manag.*, vol. 52, p. 102027, 2020.
- [20] N. Gupta and P. Sharma, "Enhancing customer targeting with artificial intelligence and big data analytics," *J. Bus. Res.*, vol. 118, pp. 134–143, 2020.
- [21] M. Sharma and V. Kapoor, "Exploring the use of artificial intelligence for segmentation in the retail industry," *Int. J. Retail Distrib. Manag.*, vol. 48, no. 10, pp. 1123–1137, 2020.
- [22] S. Huang and Y. Wang, "AI applications in marketing and consumer segmentation: A comprehensive review," *Mark. Theory*, vol. 20, no. 3, pp. 373–389, 2020.
- [23] R. Agarwal and A. Aggarwal, "Artificial intelligence in customer segmentation: Techniques and tools," *Asia Pac. J. Mark.*, vol. 31, no. 6, pp. 1057–1071, 2019.
- [24] A. Sharma and S. Desai, "Personalized marketing strategies using artificial intelligence: Case studies and industry practices," *J. Bus. Strategy*, vol. 40, no. 2, pp. 42–53, 2019.
- [25] A. Khan and S. Taneja, "AI-driven customer behavior prediction for effective segmentation," *J. Consum. Mark.*, vol. 36, no. 4, pp. 487–501, 2019.
- [26] P. Yang and M. Wang, "The role of artificial intelligence in enhancing customer engagement and segmentation," *J. Mark. Anal.*, vol. 7, no. 2, pp. 99–111, 2019.