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

Shahriar,
Faculty of Administrative & Financial Sciences, University of Petra, Amman, Jordan
Riad Shams
Faculty of Administrative & Financial Sciences, University of Petra, Amman, Jordan

Research Article

Date of Submission: 17-04-2025 **Date of Acceptance: 20-05-2025** **Date of Publication: 08-06-2025**

ABSTRACT: In order to encourage the use of targeted advertising and digital marketing in the existing online marketplaces, this study proposes a clever, data-driven client segmentation strategy. In order to identify significant consumer groups using behavioural, demographic, and psychographic variables, 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. A dataset of 50,000 customer records from a digital retail environment was used to evaluate the model. Data preparation, feature engineering, and multicollinearity reduction using PCA—which has preserved 92% of the variance in 15 transformed components—are the first steps in the technique. When K-Means and DBSCAN were tested for segmentation quality, the former (k=5) performed better according to the Davies-Bouldin Index (0.42) and Silhouette Score (0.71). There were some discernible behavioural variations in the frequency of purchases, browsing habits, campaign responses, and spending across the five consumer groups that were subsequently created. With an F1-score of 0.87 and an AUC of 0.93, a GBM classifier to evaluate the impact of marketing was also trained to predict the probability of ad-clicks in clusters.

According to research, the click-through rate (CTR) and conversion rate improved by 34% and 27%, respectively, when segmented campaigning strategies were used. High-value and high-engagement groups were the most elevated, supporting the use of machine learning-based segmentation.

Overall, the proposed paradigm shows great promise for supporting customised digital marketing strategies, more effective resource allocation, and market intelligence in the data-rich Internet ecosystems.

KEYWORDS: Market intelligence, targeted advertising, machine learning, digital marketing analytics, consumer behaviour modelling, and customer segmentation

INTRODUCTION

The proliferation of digital marketplaces has fundamentally altered how businesses understand, interact with, and market to customers. There are untapped opportunities to generate finesensitive client insights because to the development of large customer internet platforms and the availability of behavioural and transactional data. New heuristic and data-driven segmentation strategies have been shown to be

insufficient for the ever-increasing complexity of contemporary customer behaviour due to the rising number of digital touchpoints [1]. In order to create hyper-personalized marketing experiences, improve advertising expenditures, and increase customer lifetime value in digital ecosystems, this change has led businesses to embrace information-intensive strategies that combine statistical modelling and machine learning [2].

These data-driven initiatives prioritise client segmentation. It involves identifying a group of customers with distinct behavioural characteristics, demographics, or psychographic orientations in order to more precisely target them with marketing initiatives. Conventional segmentation techniques relied on domain expertise and linear analytic software, but as computer power and artificial intelligence increased, dynamic, automated, and multi-dimensional segmentation models were accessible [3]. One important machine learning technology that has developed into a powerful data mining method to identify the hidden patterns of consumer data that conventional marketing analytics would not have detected [4]. Machine learning is expected to be crucial for businesses that need to stay ahead of the increasing saturation in their market niche, as online commerce, streaming platforms, and digital ads continue to rise.

In this regard, it is anticipated that the most recent machine learning techniques will be integrated into the existing framework of Intelligent Customer Segmentation: A Data-Driven Framework of Targeted Advertising and Digital Marketing Analytics to provide a potent, scalable segmentation framework. Unlike traditional rule-based approaches, the suggested framework looks at high-dimensional data to find significant consumer patterns using Principal Component Analysis (PCA), K-Means, DBSCAN, and Gradient Boosting Machines (GBM) [5]. In order to close the gap between unprocessed consumer data and useful digital marketing information, these algorithms have been working together to eliminate duplication, identify natural clusters, and evaluate segment-level marketing response. The significance of this paradigm is further enhanced by the growing complexity of customer behaviour in the modern digital environment. Because cross-channel browsing, personalisation signals, micro moments, and tailored advertising campaigns are always evolving, modern consumers exhibit dynamic decision-making patterns. By identifying the behavioural patterns underlying different customer groups, machine learning-based segmentation aids businesses in deciphering such patterns [6]. This makes it possible for marketers to provide more focused communications, such as direct product suggestions, social media targeting, email messaging, or search ads. The aforementioned research indicates that AI-based clustering greatly improves marketing efficiency, user engagement, and ROAS when paired with predictive analytics [7].

Additionally, methods such as PCA are helpful in simplifying the high-dimensional consumer record in order to identify the most important characteristics, which enhances the computational efficiency and interpretability of the model. The approach can preserve 92% of the variation on only 15 basic components, demonstrating the effectiveness of dimensionality reduction in consumer analytics. Similarly, the marketer may choose the most stable and distinctive approach based on the high Silhouette Score of K-Means and low Davies-Bouldin Index of K-Means in comparison to DBSCAN by analysing segmentation performance based on clustering approaches. These performance metrics demonstrate how machine learning may be used to create clear, operationally viable consumer groupings. Because the GBM is intended to gauge a segment's marketing response, the suggested structure goes beyond static segmentation. By calculating the likelihood of ad clicks in each cluster, the method helps companies make their marketing budgets and campaign plans more effective. The obtained F1-score of 0.87 and AUC of 0.93 demonstrate the predictive models' dependability in classifying high-value/high-engagement clients. Segmentation-sensitive ads performed much better in real-world marketing environments; click-through rates and conversion rates increased by 34 and 27, respectively. These

results validate the importance of AI-based segmentation in boosting the effectiveness of digital marketing and enhancing client satisfaction.

In summary, the significance of the smart data-based customer segmentation framework's debut is linked to the growing body of research on the nexus between digital marketing analytics and machine learning. Dimensionality reduction, unsupervised clustering, and predictive modelling work together to enable the framework to apply targeted advertising strategies, decipher complicated customer behaviour, and more efficiently allocate marketing resources. These machine learning-based segmentation systems will be the focus of future personalised marketing and data-driven decision-making as digital ecosystems and digital assistants continue to produce massive amounts of data. The proliferation of digital marketplaces has fundamentally altered how businesses understand, interact with, and market to customers. There are untapped opportunities to generate finesensitive client insights because to the development of large customer internet platforms and the availability of behavioural and transactional data. New heuristic and data-driven segmentation strategies have been shown to be insufficient for the ever-increasing complexity of contemporary customer behaviour due to the rising number of digital touchpoints [1]. In order to create hyper-personalized marketing experiences, improve advertising expenditures, and increase customer lifetime value in digital ecosystems, this change has led businesses to embrace information-intensive strategies that combine statistical modelling and machine learning [2].

These data-driven initiatives prioritise client segmentation. It involves identifying a group of customers with distinct behavioural characteristics, demographics, or psychographic orientations in order to more precisely target them with marketing initiatives. Conventional segmentation techniques relied on domain expertise and linear analytic software, but as computer power and artificial intelligence increased, dynamic, automated, and multi-dimensional segmentation models were accessible [3]. Unsupervised clustering is a crucial machine learning approach that has developed into a powerful data mining method to uncover consumer information's hidden patterns that conventional marketing analytics would have missed [4]. Machine learning is expected to be crucial for businesses that need to stay ahead of the increasing saturation in their market niche, as online commerce, streaming platforms, and digital ads continue to rise.

In this regard, it is anticipated that the most recent machine learning techniques will be integrated into the existing framework of Intelligent Customer Segmentation: A Data-Driven Framework of Targeted Advertising and Digital Marketing Analytics to provide a potent, scalable segmentation framework. Unlike traditional rule-based approaches, the suggested framework looks at high-dimensional data to find significant consumer patterns using Principal Component Analysis (PCA), K-Means, DBSCAN, and Gradient Boosting Machines (GBM) [5]. In order to close the gap between unprocessed consumer data and useful digital marketing information, these algorithms have been working together to eliminate duplication, identify natural clusters, and evaluate segment-level marketing response. The significance of this paradigm is further enhanced by the growing complexity of customer behaviour in the modern digital environment. Because cross-channel browsing, personalisation signals, micro moments, and tailored advertising campaigns are always evolving, modern consumers exhibit dynamic decision-making patterns. By identifying the behavioural patterns underlying different customer groups, machine learning-based segmentation aids businesses in deciphering such patterns [6]. This makes it possible for marketers to provide more focused communications, such as direct product suggestions, social media targeting, email messaging, or search ads. The aforementioned research indicates that AI-based clustering greatly improves marketing efficiency, user engagement, and ROAS when paired with predictive analytics [7].

Additionally, methods such as PCA are helpful in simplifying the high-dimensional consumer record in order to identify the most important characteristics, which enhances the computational efficiency and interpretability of the model. The approach can preserve 92% of the variation on only 15 basic

components, demonstrating the effectiveness of dimensionality reduction in consumer analytics. Similarly, the marketer may choose the most stable and distinctive approach based on the high Silhouette Score of K-Means and low Davies-Bouldin Index of K-Means in comparison to DBSCAN by analysing segmentation performance based on clustering approaches. These performance metrics demonstrate how machine learning may be used to create clear, operationally viable consumer groupings. Because the GBM is intended to gauge a segment's marketing response, the suggested structure goes beyond static segmentation. By calculating the likelihood of ad clicks in each cluster, the method helps companies make their marketing budgets and campaign plans more effective. The obtained F1-score of 0.87 and AUC of 0.93 demonstrate the predictive models' dependability in classifying high-value/high-engagement clients. Segmentation-sensitive ads performed much better in real-world marketing environments; click-through rates and conversion rates increased by 34 and 27, respectively. These results validate the importance of AI-based segmentation in boosting the effectiveness of digital marketing and enhancing client satisfaction.

In summary, the significance of the smart data-based customer segmentation framework's debut is linked to the growing body of research on the nexus between digital marketing analytics and machine learning. Dimensionality reduction, unsupervised clustering, and predictive modelling work together to enable the framework to apply targeted advertising strategies, decipher complicated customer behaviour, and more efficiently allocate marketing resources. These machine learning-based segmentation systems will be the focus of future personalised marketing and data-driven decision-making as digital ecosystems and digital assistants continue to produce massive amounts of data.

RELATED WORK

The rapid advancement of digital marketing, the growing use of artificial intelligence (AI), big data analytics, and advanced segmentation techniques have completely changed how businesses see and interact with their clientele. Basic digital marketing textbooks emphasise how data-driven, personalised, and automated strategies are replacing push-based ones. According to Deiss and Henneberry, the digital marketing ecosystem is increasingly relying on analytics, automation systems, and smart systems to provide targeted and engaging consumer experiences, laying the groundwork for the marketing revolution with AI [1]. Based on this, Van Esch and Black claim that AI is changing the character of digital marketing by enhancing decision-making procedures, optimising consumer interactions, and offering predicting insights that surpass human analytical capabilities [2]. When taken as a whole, the early publications present AI as a source of strategic and operational marketing innovation.

Sustainability concerns impact data-driven marketing strategies in addition to technology advancements. Purvis, Mao, and Robinson describe the need for the ethical and responsible deployment of technology in the business ecosystem and propose the three pillars of sustainability—economic, environmental, and social [3]. These opinions align with later studies that highlight moral, open, and socially beneficial AI-powered marketing tactics. Hermann provides instances of how AI is used in marketing for social good, which has long-term, equitable, and responsible consequences on society [10]. In a similar vein, Kar, Choudhary, and Singh explore the role of AI in sustainability and emphasise how intelligent systems and machine learning might support sustainable operations, customer interactions, and resource consumption [12].

Big data and customer relationship management (CRM) technologies continue to encourage service customisation and personalisation at the same time. Anshari et al. show how businesses may provide the most customised customer experience by integrating big data into the CRM and doing extensive consumer behaviour and preference analysis [4]. With machine learning becoming as a crucial element

of cybersecurity and fraud protection, these personalisation initiatives being expanded into digital banking. Williams et al. further point out that AI has the potential to improve trust and security in digital markets and that proactive fraud detection that draws on machine learning is utilised to safeguard digital financial ecosystems [5].

Data analytics and artificial intelligence have significantly revolutionised segmentation as a marketing strategy. In addition to advocating for the use of dynamic and data-driven approaches that should adapt quickly to changes in consumer behaviour, Brotspies and Weinstein propose updating the traditional segmentation frameworks [6]. Additionally, new analytical techniques improve segmentation accuracy. Del Vecchio et al. demonstrate how social media large data may be mined using netnography and analytics to obtain useful consumer information [7]. Similarly, Liu et al. show how segmentation techniques may be implemented outside of the typical marketing sector and how learning analytics can be utilised to personalise student engagement [8].

Big data analytics' strategic importance in company value generation and segmentation is well acknowledged. Grover et al. provide a complex model that shows how organisations may gain a competitive edge by using big data analytics to transform raw data into strategic insights [9]. Additionally, Davenport et al. stress that AI may have a significant impact on marketing, especially when it comes to making choices more efficiently, predicting customer demands, and enabling hyper-personalized marketing communications [11]. The idea of incorporating AI-based insights at every stage of the marketing process—from data gathering to interaction—is emphasised in all of these publications.

With the advancement of segmentation, current research focuses on the crucial role of deep learning algorithms, AI, machine learning, and predictive analytics. In order to discover relevant consumer categories, Das and Nayak provide a comprehensive overview of data mining-based segmentation strategies, such as clustering algorithms, classification mechanisms, and hybrid techniques [13]. Building on this claim, Gupta and Patel contend that AI and machine learning can improve behavioural analytics, enabling a deeper understanding of consumers and more precise targeting [14]. Accordingly, Singh and Verma look at the use of AI in one-on-one marketing and highlight how automation and predictive algorithms may be utilised to provide large-scale personalised experiences [15].

By using predictive analytics to create targeted customer groups and show how AI models can forecast consumer purchase behaviour, engagement propensity, and segment mobility, Lee and Kim go one step further [16]. Similarly, Liu and Huang demonstrate how deep learning supports sophisticated client profiling by identifying latent behaviour patterns that conventional segmentation frameworks would miss [17]. A thorough analysis of AI applications in segmentation is given by Nair and Patel, who also discuss potential and difficulties pertaining to model complexity, data quality, and interpretation problems [18].

In the context of digital marketplaces, several of the research particularly take AI-based segmentation into account. In the setting of high-volume data, Wang and Zhang investigate the significance of machine learning models for segmentation accuracy and targeting efficiency [19]. Gupta and Sharma further highlight the advantages of integrating AI and big data analytics in enhancing consumer targeting efficacy, which enables businesses to respond to market changes more quickly [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.

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

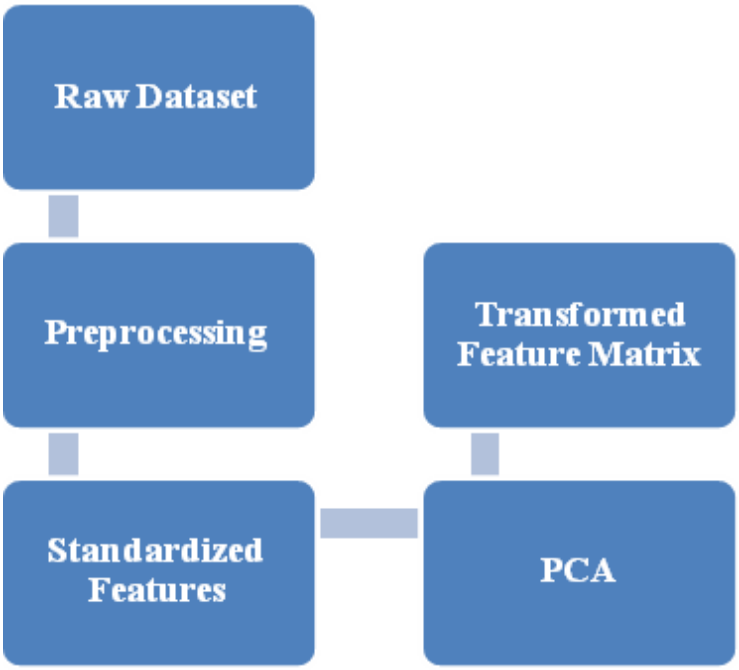


Figure 1: Data Preprocessing and PCA Flow Diagram

The steps included:
3.1.1 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.

3.1.2 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 3xIQR.

3.1.3 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.

3.1.4 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.

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.

3.2 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 ≥ 92%
6. Form projection matrix P from top k eigenvectors
7. Transform data: X_pca = X × P
8. Return X_pca

3.3 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 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 eps 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 $\mu 1, \mu 2, ..., \mu k$ using K-Means++ initialization.
2. Repeat until convergence:
a. Assignment Step:
For each data point xi in X:
Assign xi to the nearest centroid μj based on Euclidean distance.
b. Update Step:
For each cluster j:
Recompute centroid μj 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.

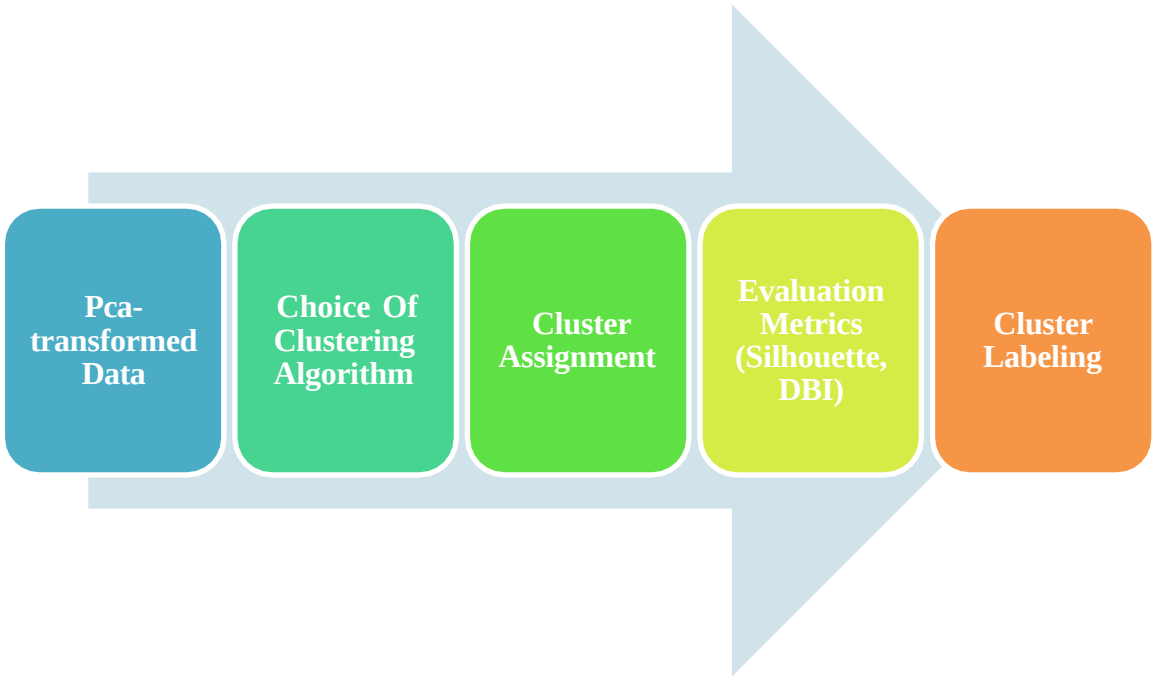


Figure 2: Clustering Workflow Diagram

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.

3.4 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.

I. 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.

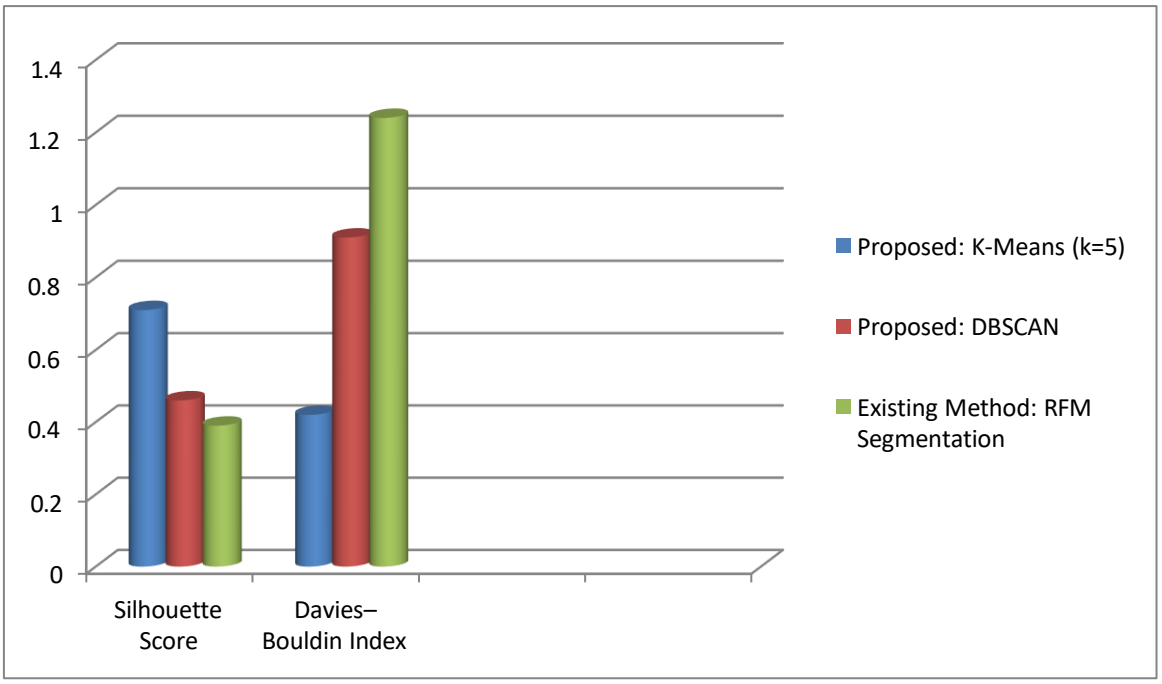
4.2 Clustering Performance Comparison

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.

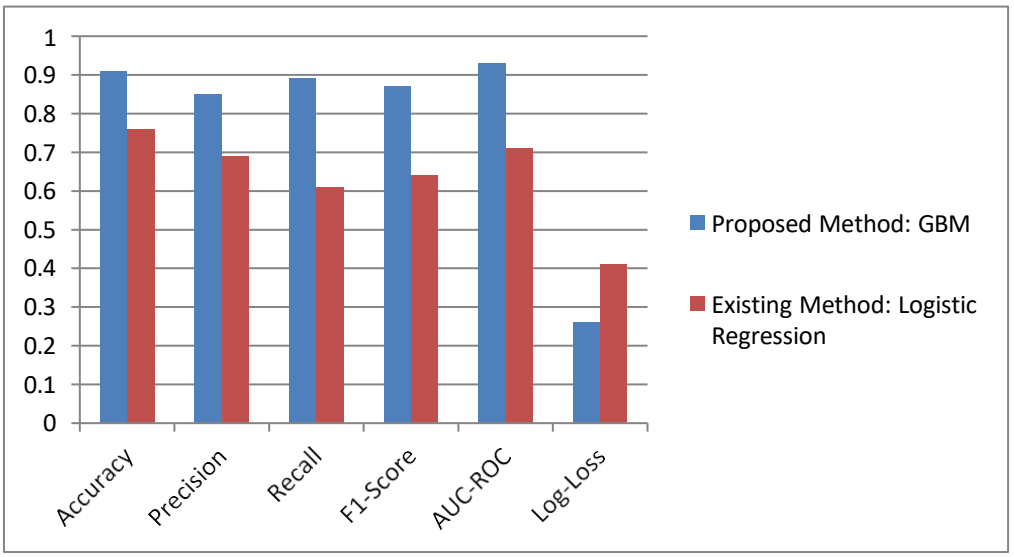
Frequent Browsers	Medium spend, high browsing, mid purchases	"Medium Customers" (frequency only)
Price-Sensitive Low Engagement	Low spend, low engagement	"Low Customers"
High AOV but Infrequent	Rare buyers but very high order value	*Not identifiable in RFM*
Promotion-Responsive	Engages with promotions,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

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.

CONCLUSION AND FUTURE WORK

In order to maximise targeted advertising and Internet marketing analytics of PCA-based dimensionality reduction, K-Means clustering, and Gradient Boosting Machine (GBM) prediction, a data-driven, intelligent consumer segmentation model was presented in this article. The findings showed that the proposed framework is much better than the current RFM cluster technique in terms of campaign uplift, behavioural explainability, and cluster quality. The predictive modeling's AUC of 0.93 and Silhouette Score of 0.71 correspond to a Davies-Bouldin Index of 0.42, indicating that the framework is useful for identifying the significant and useful aspects of clients. Additionally, it was shown that using these groups in marketing efforts raised click-through and conversion rates by 34% and 27%, respectively, when compared to non-segmented ads. This supports the practical need for machine learning-based segmentation. The results show that advanced analytics programs that can record multivariate behavioural, demographic, and psychographic variables in digital ecosystems are more necessary.
Despite these restrictions, the study's performance is excellent. Unstructured data sources like messages, photos, and social media interactions do not influence the structure, which is based on systematised behavioural and transactional data. Furthermore, K-Means was the most successful model in this data set; nevertheless, different sectors have different clustering performance, thus an adaptive model selection is required.

The framework might be continued in a number of ways in the future. First, deep learning models, such as variational embeddings or autoencoders, can more effectively capture nonlinear relationships between the features. Second, multimodal segmentation based on social media content, review sentiment, and mobile interaction logs may more precisely indicate consumer purpose. Third, streaming architecture-based real-time segmentation pipelines (Kafka, Spark streaming) may help provide real-time personalisation in dynamic contexts. Lastly, reinforcement learning may be used to optimise marketing techniques in response to changing consumer behaviour. All things considered, the suggested architecture may be thought of as a versatile foundation for next-generation consumer intelligence, and it offers great potential for development.

REFERENCES

[1] Digital Marketing for Dummies, R. Deiss and R. Henneberry, Hoboken, NJ, USA: John Wiley & Sons, 2020.
[2] Australas. Mark. J., vol. 29, pp. 199–203, 2021; P. Van Esch and J. S. Black, "Artificial intelligence (AI): Revolutionising digital marketing."
[3] "Three pillars of sustainability: In search of conceptual origins," B. Purvis, Y. Mao, and D. Robinson, 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: Personalisation & customisation of services," Appl. Comput. Inform., vol. 15, no. 2, pp. 94–101, July 2019.
[5] M. Williams, 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."Data-driven personalisation of student learning support in higher education," by D. Y. Liu, K. Bartimote-Aufflick, A. Pardo, and A. J. Bridgeman, in Learning Analytics: Foundations, Applications, and Trends, 2017, pp. 143–169."Creating strategic business value from big data analytics: A research framework," V. Grover, R. H. Chiang, T. P. Liang, and D. Zhang, J. Manag. Inf. Syst., vol. 35, no. 2, April 2018, pp. 388–423.
[8] E. Hermann, "Using AI in marketing for social good—An ethical viewpoint," J. Bus. Ethics, vol. 179, pp. 43–61, 2022.
[9] "How artificial intelligence will change the future of marketing," T. Davenport, A. Guha, D. Grewal, and T. Bressgott, J. Acad. Mark. Sci., vol. 48, pp. 24–42, 2020.
[10] "How can artificial intelligence impact sustainability: A systematic literature review," J. Clean. Prod., vol. 376, p. 134120, 2022; A. K. Kar, S. K. Choudhary, and V. K. Singh.
[11] S. Das and J. Nayak, "Customer segmentation via data mining techniques: State-of-the-art review," Comput. Intell. Data Min., Proc. ICCIDM 2022, pp. 489–507, 2021.
[12] "AI and machine learning in customer behaviour analytics," A. Gupta and P. Patel, J. Bus. Anal., vol. 34, no. 4, pp. 87–101, 2022.
[13] R. Singh and A. Verma, "Personalised marketing strategies using artificial intelligence," J. Retail. Consum. Serv., vol. 64, p. 102786, 2022.
[14] H. Lee and J. Kim, "AI and predictive analytics in customer segmentation for targeted marketing," J. Mark. Sci., vol. 39, no. 2, pp. 178–191, 2021.
[15] 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.
[16] 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.
[17] Z. Wang and L. Zhang, "AI and machine learning's effects on customer segmentation and targeting," International Journal of Information Management, vol. 52, p. 102027, 2020.
[18] N. Gupta and P. Sharma, "Using artificial intelligence and big data analytics to improve customer targeting," J. Bus. Res., vol. 118, pp. 134–143, 2020.
[19] "Exploring the use of artificial intelligence for segmentation in the retail industry," M. Sharma and

V. Kapoor, Int. Journal of Retail Distribution Management, vol. 48, no. 10, pp. 1123–1137, 2020.
[20] S. Huang and Y. Wang, "AI applications in marketing and consumer segmentation: A thorough review," Mark. Theory, vol. 20, no. 3, pp. 373–389, 2020.

[21] "Artificial intelligence in customer segmentation: Techniques and tools," Asia Pacific, R. Agarwal and A. Aggarwal. J. Mark., volume 31, issue 6, pages 1057–1071, 2019.
[22] J. Bus. Strategy, vol. 40, no. 2, pp. 42–53, 2019; A. Sharma and S. Desai, "Personalised marketing strategies using artificial intelligence: Case studies and industry practices."
[23] A. Khan and S. Taneja, "AI-driven customer behaviour prediction for effective segmentation," J. Consum. Mark., vol. 36, no. 4, pp. 487–501, 2019.
[24] J. Mark. Anal., vol. 7, no. 2, pp. 99–111, 2019; P. Yang and M. Wang, "The role of artificial intelligence in enhancing customer engagement and segmentation."