The American Journal of Engineering and Technology
ISSN 2689-0984 Volume 08 - 2026

An Explainable AI Framework for Marketing Customer Segmentation and Campaign Response Prediction

Pritesh Gehlot
Designation Sr Architect, USA

Soumyajit Mukherjee
Designation Sr Manager, Data, Analytics and AI Engineering Technology, USA

Received: 14 Oct 2026 | Received Revised Version: 23 Oct 2026 | Accepted: 04 Nov 2026 | Published: 12 Nov 2026

Volume 08 Issue 11 2026 | Crossref DOI: 10.37547/tajet/Volume08Issue11-04

Abstract

The segmentation of clients and the forecast of marketing campaign response are essential for optimising marketing expenditures and enhancing customer participation. This work suggests an XAI approach for segmenting the clients, predicting the responses using machine learning, and targeting them based on profit. Specifically, 2,055 cleansed customer records were extracted from the Kaggle Customer Personality Analysis dataset through preprocessing. We combine multiple clustering methods using consensus clustering to find the different segments of consumers, using Principal Component Analysis (PCA). The performance of seven machine learning classifiers is assessed using five-fold cross-validation, considering several performance metrics. The best accuracy (ACC) is 87.59% for Gradient Boosting (GB) and the ROC-AUC is 0.909 for Logistic Regression. The segment integration yields an improved recall (REC) rate of 79.4% and F1 score (F1) of 0.649, compared to the previous situation of REC value 73.0% and F1 value 0.622. The resulting profit-optimised solution is a projected profit of 61.1 currency units per 100 consumers. Moreover, elements such as catalogue purchases, recency, and campaign history are identified as elements that are significant via the techniques of SHapley Additive ExPlanations (SHAP) and permutation feature significance. The proposed approach will enable more accurate client segmentation, targeted marketing campaigns, and efficient use of resources. More testing is needed on a variety of real-life datasets to ensure greater generalisability, but the outcomes are encouraging.

Keywords: Customer segmentation, explainable artificial intelligence, marketing campaign prediction, machine learning, consensus clustering.

© 2026 Gehlot, P, & Mukherjee, S. This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). The authors retain copyright and allow others to share, adapt, or redistribute the work with proper attribution.

Cite This Article: Gehlot, P, & Mukherjee, S. (2026). An Explainable AI Framework for Marketing Customer Segmentation and Campaign Response Prediction. The American Journal of Engineering and Technology, 8(11), 35–58. Retrieved from https://theamericanjournals.com/index.php/tajet/article/view/8606

1. Introduction

Customer Segmentation plays a pivotal role for modern businesses to grasp and cater to the varying requirements and inclinations of customers. Segmentation enables more effective marketing targeting, thereby improving customer satisfaction, customer loyalty and customer retention [1]. This is done by segmenting the customers into various groups with similar attributes such as demographics, behavioral traits, buying trends, etc. Our clients' groups are known to us and we can develop individual marketing strategies. This means your customers are happier and you make more sales. Organisations do this by analyzing a broad range of data from various sources, including demographic data, behavioral data, and purchasing patterns, which helps them distinguish between different customer segments and identify patterns [2]. This segmentation is then used to inform custom marketing tactics, elevating the effectiveness of marketing efforts and contributing to business growth.

Separating a target audience into subsets defined by shared traits, wants, or actions is known as customer segmentation

and is an essential marketing tactic [3]. Businesses can better meet customer needs, allocate their promotion budget wisely and develop targeted marketing strategies with proper segmentation [4]. Artificial intelligence provides sophisticated ways to analyse consumer data and spot trends that could be hard to find using more traditional methods. Machine learning (ML) algorithms can segment customers into different groups based on their engagement, purchases, website activity, and transaction history. Recommending different products and services relevant to a particular customer and predicting its future buying habit using a recommender system and predictive model.

Traditional marketing strategies often target broad customer groups, and they do not consider individual consumer buying habits, responsiveness, or the potential profitability of the marketing campaign. The interpretation of models is limited and the issue of class imbalance is already a challenge when predicting the correct answer. Therefore, a cohesive strategy that incorporates profit-driven threshold determination, ML, explainable AI, and customer segmentation is needed to identify these individuals, understand why they act in certain ways, prevent marketing waste, and improve marketing effectiveness. The following are key findings of the study:

- The data is then used to develop a consensus clustering approach which sorts customers into a series of customer segments based on their demographics, buying habits, and RFM characteristics.
- Tests seven ML classifiers with a variety of performance measures and five-fold cross-validation.
- Compared to a baseline model, it seeks to determine if customer segment information enhances response identification.
- Determines the most important parameters impacting campaign response forecasts by using SHAP and permutation feature significance.
- Increases the efficiency of marketing budget distribution by optimising the campaign decision threshold, which in turn prioritises possible responders.
- Uses segmented client reaction probability and profitability to inform targeted marketing strategies.

A. Organization of the Paper

The following is the outline of the paper: The relevant literature and research gap are reviewed in Section II. Models and technique are detailed in Section III. The results of the experiments and an analysis of the performance are presented in Section IV. Lastly, the work is concluded and future research goals are outlined in Section V.

2. Literature Review

ML, DL, and hybrid techniques are increasingly being used in marketing optimisation, campaign response prediction, and consumer segmentation, according to the available literature. S. Trivedi et al. (2026) The model successfully learns patterns based on both time and structure by integrating LSTM, XGBoost, and an attention-based fusion mechanism. The suggested hybrid model achieves better results than the baseline models in several respects, including ACC (96.8%), PRE (95.9%), REC (97.4%), and F1 (96.6%). In contrast, the ACC rates of the baseline models were 85.2%, 90.4%, XGBoost, 93.6%, and 94.8%, respectively [5].

A. A. Durib and A. Gökmen (2026) include data loading and exploratory analysis, as well as SMOTE pretreatment for unbalanced data. With a ROC-AUC of 97.2% and an ACC of 92.9%, XGBoost performed admirably on the EdTech Leads dataset. In comparison to the other models, LightGBM trained faster and had a ROC-AUC of 92.7% and an ACC of 90.5% on the unbalanced Bank Marketing dataset. Tags, Lead Profile, Engagement Score, and Average Duration were found by SHAP analysis. One of the most important factors that predicted [6].

S.Sowndharya and R.Hariharan (2026) Combine the advantages of DL and reinforcement learning. RF (ACC: 88.2%, F1: 87.4%) and GB (ACC: 90.1%, F1: 89.3%), two classic AI models, and traditional segmentation-based techniques a were outperformed by ADRPA on a large-scale digital marketing dataset. The AUC (Area Under Curve) value of 0.96 indicates that ADRPA obtained higher ACC in classification and improved capacity to forecast the customer conversion behaviour. The model earned a REC of 92.7%, a PRE of 93.5%, and an F1 of 93.1% [7].

A. Khowaja et al. (2025) One of the most important marketing strategies, customer segmentation enables organisations to personalise offerings and enhance customer engagement. This article demonstrates a customer segmentation strategy that makes use of ML in conjunction with if-else statements in order to provide tailored offers to

various client coalitions. For more accurate consumer segmentation, leading to more relevant ads and interactions, a hybrid model combining K-means clustering with a rule-based system is employed [8].

M. El-Hajj and M. Pavlova et al. (2024) create a DT prediction model to find out what factors are affecting consumer behaviour and how to better target campaigns. Starting with an initial ACC of 87.3%, the technique entails developing the DT model. But because of the imbalance in the classes, the model had trouble with REC and ACC. This was addressed by using a resampling strategy, which greatly enhanced the model's performance, raising the F1 from 49% to 74.2% and REC from 44% to 83.1% [9].

S. Agarwal (2024) uses the Mall Customers dataset to examine how k-means clustering may be used for customer segmentation based on demographics and shopping behaviour, demonstrating how ML can do this. Following a comprehensive literature review, the paper implements the clustering technique, evaluates models, and discusses the findings. With a Silhouette score of 0.69777 and a Davies-Bouldin Index (DBI) of 0.42110, we can confirm that the model is accurate [10].

N. Godcares et al. (2023) Economists may learn more about their consumers by customer segmentation, which involves dividing them into subsets defined by shared traits, interests, and other factors. A marketing strategy suggestion model was constructed using the generated clusters. Which channel with the most demands was also shown by the clustering result. The BIRCH clustering technique we utilised yielded a performance score of 75.1% for silhouettes and 178.7% for Calinski-Harasbaz scores [11]. There has been little research that integrates explainable AI with profit-based targeting; most studies only deal with client segmentation or campaign response prediction separately. Additionally, there has been inadequate attention to class imbalance, segment-wise profitability, and interpretability at the consumer level. Hence, to enhance targeted advertising and resource distribution, a unified framework is required that integrates consensus clustering, ML, SHAP-based explanations, and profit-based threshold optimisation.

3. Methodology

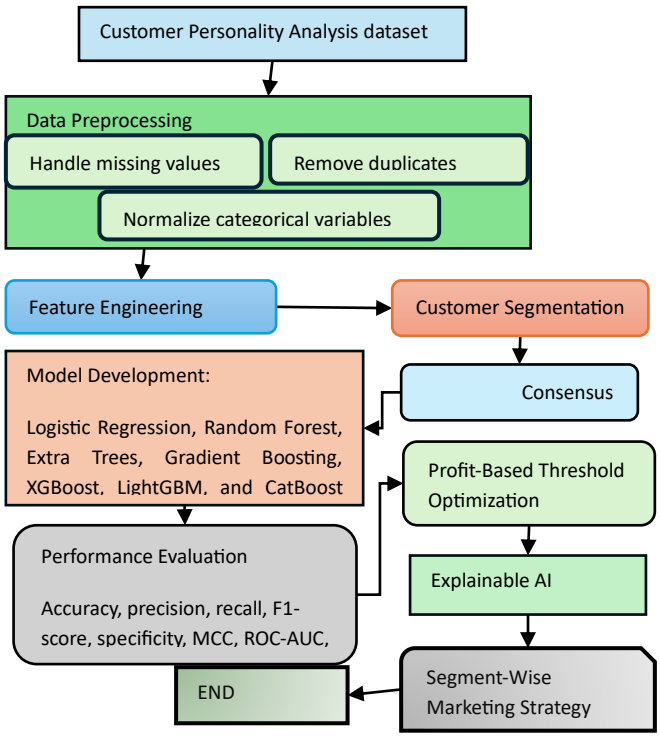


Fig. 1. Proposed Framework for Customer Segmentation and Campaign Prediction

Customer segmentation using principal component analysis (PCA) with various clustering methods and consensus clustering are all part of the technique. Other steps include feature engineering and RFM analysis. In order to train and evaluate seven ML classifiers using conventional performance indicators, five-fold cross validation is employed. We find the

best campaign targeting tactics for optimising the profit threshold, and we employ SHAP and permutation significance to explain the prediction variables. Finally, segment-specific response and profit rates facilitate targeted marketing while keeping campaign expenditures to a minimum. Figure 1 shows the data flow diagram that was created throughout the research design phase. The proposed methodology and its implementation process are explained in the following steps:

A. Data Source and Pre-processing

The Kaggle Customer Personality Analysis dataset [12] has 29 characteristics and 2,240 observations. There were two constant columns, 182 duplicate profiles, and 24 missing income figures discovered during pre-processing. The process involves removing duplicate entries and three records with extremely unlikely ages, standardizing the categories labels, and clipping negative expenditure figures to zero. The missing income values are filled in using training-fitted KNN, while the remaining values are filled in using the medians or modes of the training set. A training-derived IQR fence is used to identify income outliers, and a training-derived reference date is used to build features for age and customer tenure. Last but not least, the cleaned data set had 2055 rows and 32 columns with full data and was stratified-sampled to split the data into 80% training (1644 rows) and 20% testing (411 rows).

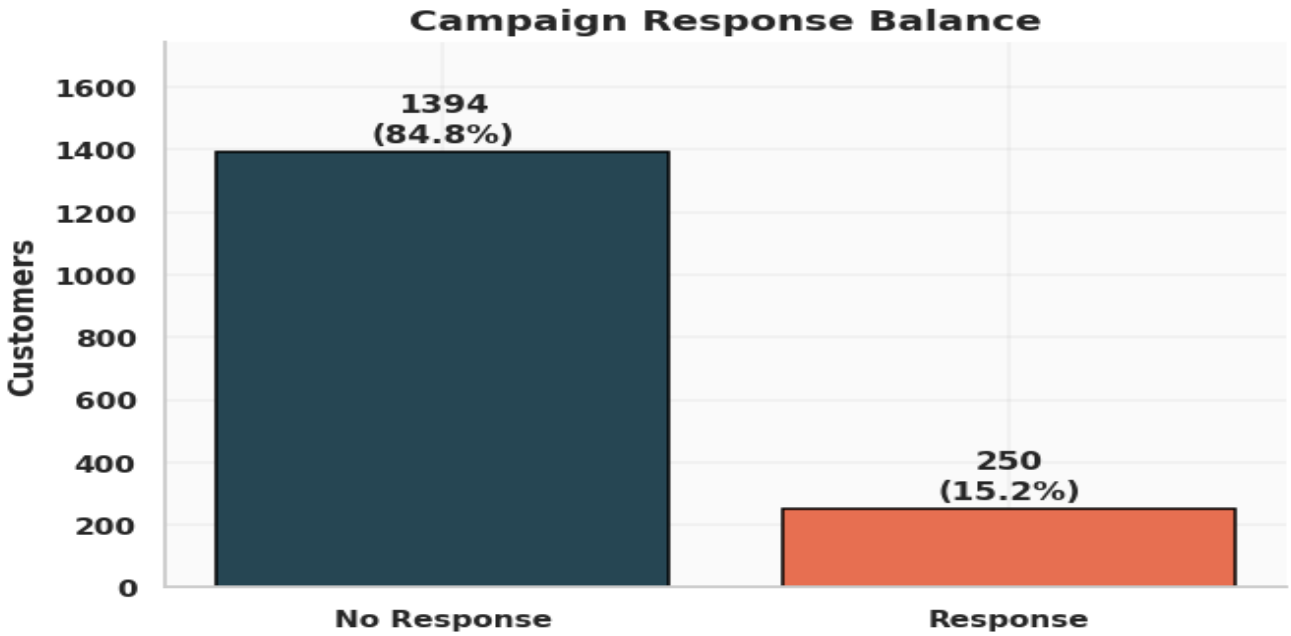


Fig. 2. Class Distribution of Marketing Campaign Responses

Figure 2 shows how the 1,644 consumers who participated in the campaign were distributed throughout the dataset. Among them, 1,394 (84.8%) chose not to participate in the campaign, while 250 (15.2%) participated.

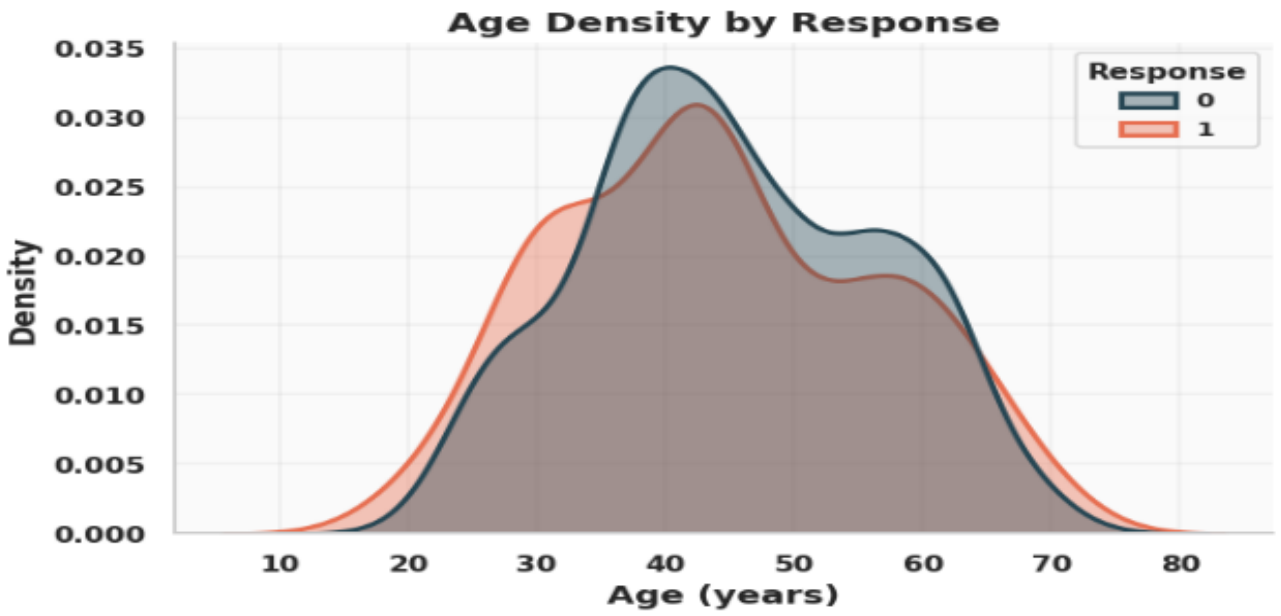


Fig. 3. **Age Distribution of Customers by Marketing Campaign Response**

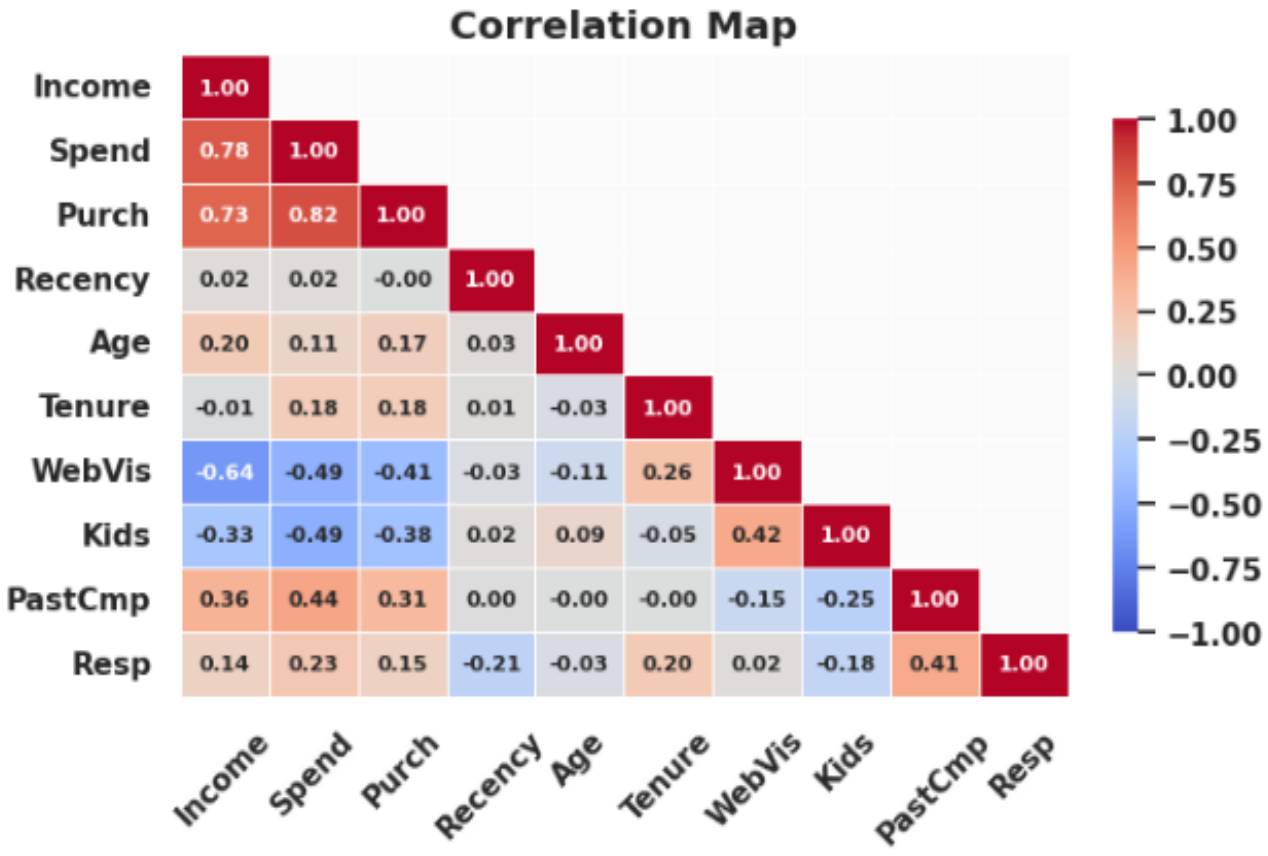


Fig. 4. **Correlation Heatmap**

Figure 3 shows the distributions of ages for both the campaign's responders and non-responders. The two groups' age distributions are very similar, with the peak density occurring between the ages of 40 and 45. There is a lot of overlap in the distributions of responders across most age categories, although there is a comparatively higher density among younger clients, especially those in the 30-35 age range.The relationships among client demographics, purchasing behaviour, website visits, prior campaign acceptability, and the current campaign reaction are visually represented in Figure 4. There is a high degree of association between income and both expenditure and purchases (0.82), and there is also a substantial positive correlation between income and purchases (0.73). The present answer is most positively correlated with previous campaign acceptability (0.41), while recency is negatively correlated (−0.21). Compare the 5 RFM Segments in Figure 5. The 374 responses to the Champion survey had the highest response rate of 29.9%, followed by the Potential group at 16.5%, the Loyal group at 15.0%, the At Risk group at 7.8%, and the Hibernating group at 2.2%.

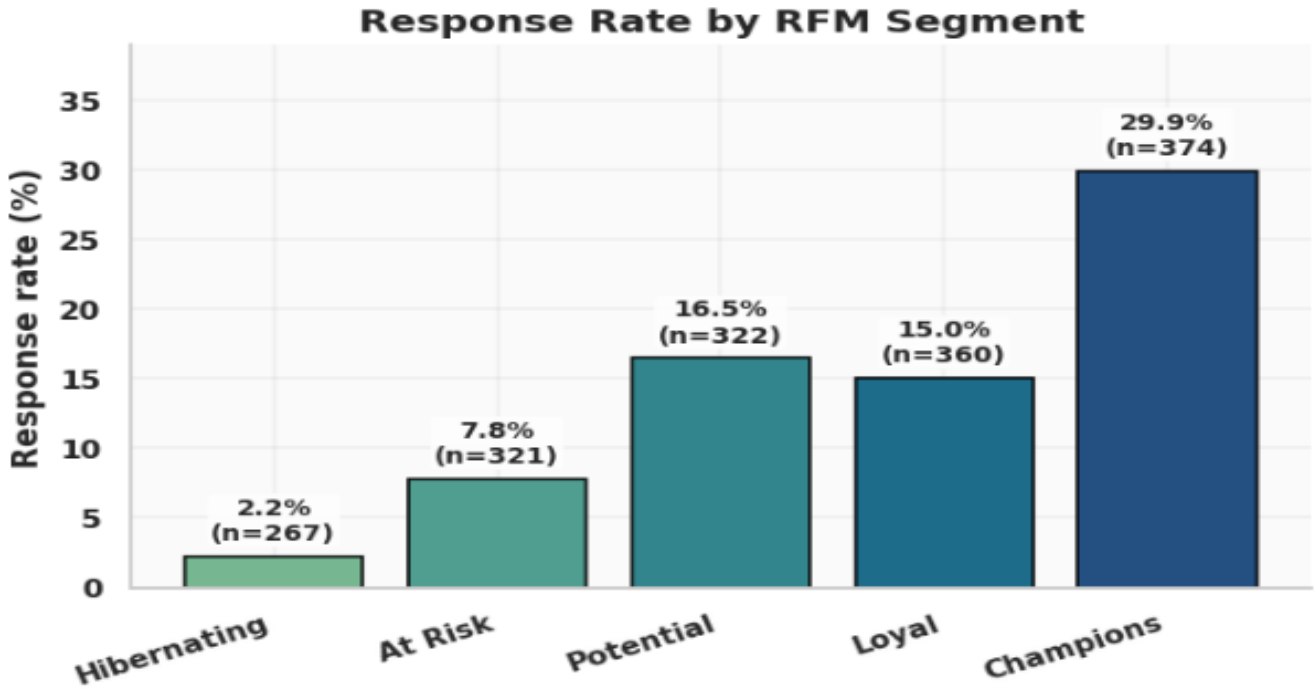


Fig. 5. Marketing Campaign Response Rate Across RFM Customer Segments

B. Feature Engineering

Feature engineering added additional attributes to the cleaned data, which included customer tenure, family size, total spending, purchase frequency, campaign acceptance, average spending per purchase, deal sensitivity, premium-product ratio, web conversion, and product spending shares, increasing their number from 32 to 52. The Shannon entropy can be used to measure the variability of the behaviour for purchasing channels and product categories. Based on their RFM ratings, the consumers are segmented into 5 categories – Hibernating, At Risk, Potential, Loyal, and Champions. These categories are defined using quantile limits from the training set. Customer demographics, purchasing habits, campaign responses, and behavioural variety are all stored in these manufactured characteristics, which are then used for segmenting customers and predicting campaign responses. In the finished dataset, there are no blanks.

C. Customer Segmentation and Optimal Cluster Selection

Using K-Means, GMM, and Agglomerative Clustering in PCA space, the number of clusters is examined from (k=2) to (k=8). The following measures are used to assess cluster quality: Silhouette score, Davies-Bouldin index, Calinski-Harabasz score, and GMM BIC. To measure cluster stability, we use repeated 80% subsampling and the Adjusted Rand

Index (ARI). All three normalised silhouette scores—Calinski-Harabasz, inverse Davies-Bouldin, and stability—are averaged to get a composite score. The use of silhouette, elbow, BIC, stability, and composite score plots allows this to be optimised for $k \geq 3$.

D. Consensus-Based Customer Segmentation

The co-association matrix of ten K-Means runs, GMM, Agglomerative, Birch, and Spectral Clustering is fused to provide consensus segments. The Davies-Bouldin index, silhouette score, Calinski-Harabasz score, and ARI are used to evaluate the segmentation quality. The training-set spending forms a hierarchy for consumers, with S0 being the lowest and S2 the highest. The test customers are distributed based on the centroid that is nearest to them, as determined by the training-set. The dataset contains 1,644 training records and 411 test records. The test set comprises 157 customers in S0, 124 in S1, and 130 in S2.

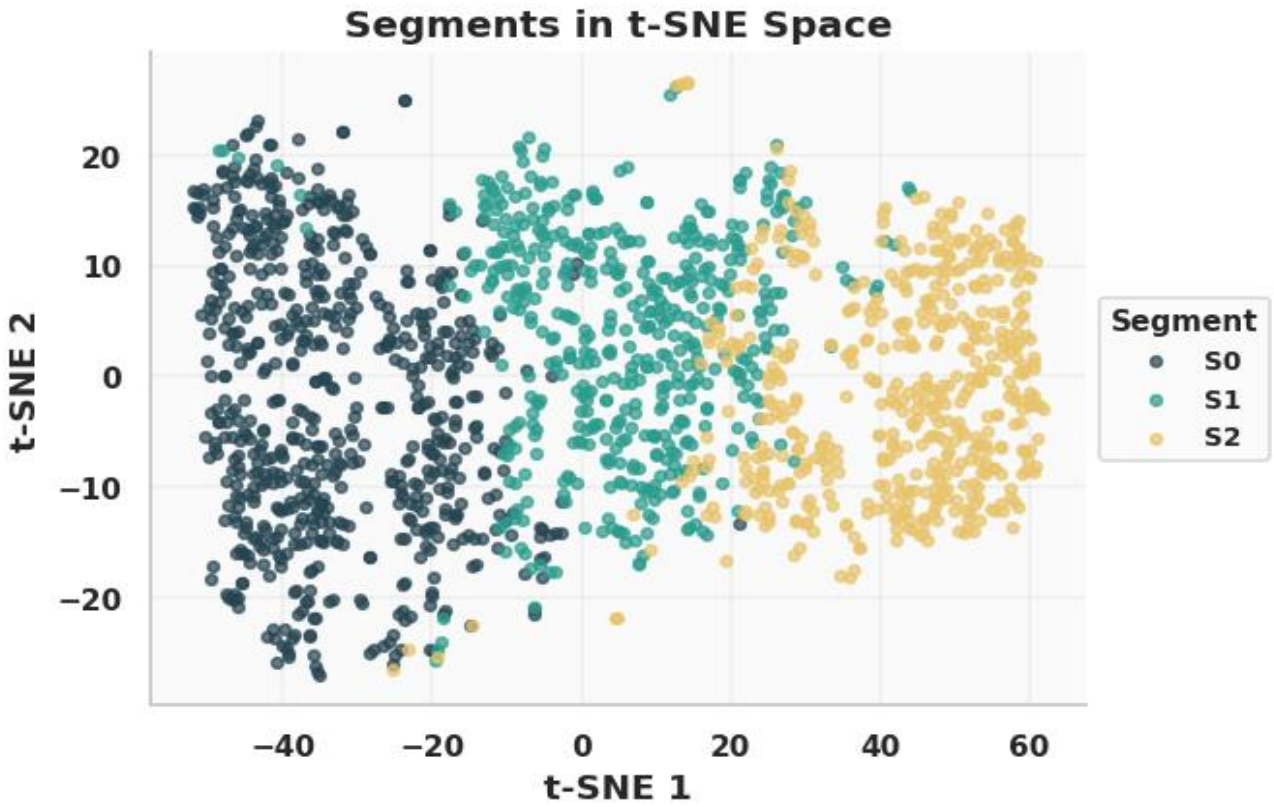


Fig. 6. Visualization of Customer Segments in t-SNE Space

Figure 6 shows the results of t-distributed Stochastic Neighbour Embedding (t-SNE) on the distribution of three client segments: S0, S1, and S2. In terms of distribution, S0 is mostly on the left, S1 is in the middle, and S2 is mostly on the right. While there is considerable overlap between nearby segments, the visualisation clearly shows that there is significant difference across consumer groups with different behavioural traits.

E. Customer Segment Profiling and Behavioral Comparison

Figure 7 shows three customer categories (S0, S1, and S2) based on normalised values for income, expenditure, recency, purchases, website visits, and number of children. Class S0 are bargain hunters, frugal families, Class S1 are traditional families who pay full price and Class S2 are premium, childless consumers. The radar map also identifies differences in client demographics and buying habits to support segment-specific marketing strategies.

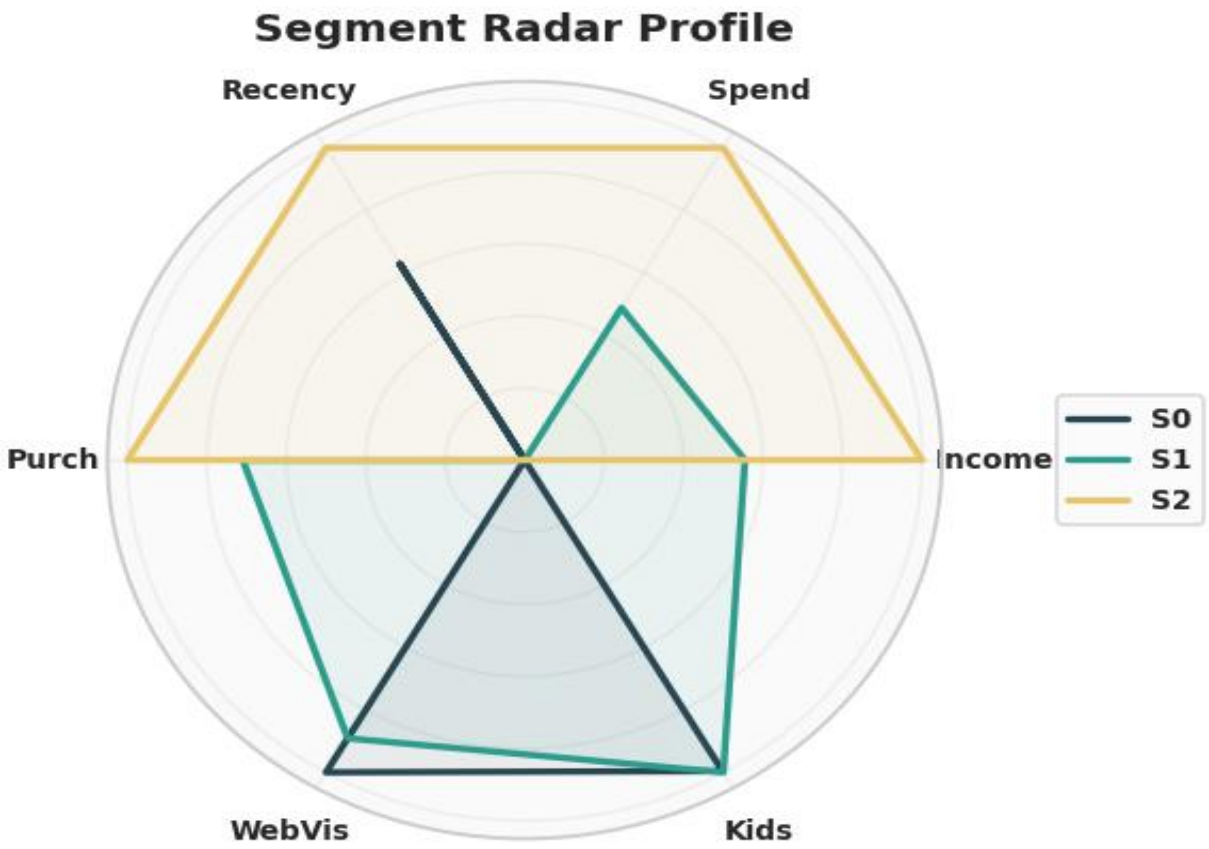


Fig. 7. Radar-Based Comparison of Customer Segment Profiles

F. Machine Learning Models and Hyperparameter Configuration

The marketing campaign responses are predicted using seven ML classifiers that use CV 5-fold.

1. Logistic regression (LogReg)

A pipeline is used to apply logistic regression, with PowerTransformer serving as the feature transformation. The model's regularisation parameter (C) was set to 0.5 during convergence, class weight was set to "balanced," and max iter was set to 3000 in order to handle the class imbalance.

Random Forest (RF)The parameters for RF are as follows: n estimators = 400, min samples_leaf = 2, and class weight = 'balanced subsample.' Each leaf will contain at least two samples. To enable parallel processing, set n_jobs = -1; for reproducibility, fix the random seed.

2. Extra Trees (ET)

Extra Trees are constructed from 400 randomised decision trees, with min_samples_leaf = 2 and class_weight = 'balanced_subsample'. The use of parallel processing with a fixed seed improves computing efficiency and repeatability simultaneously.

3. Gradient Boosting (GB)

The default estimator and learning setup from Scikit-learn are used with random_state = SEED to execute GB and achieve repeatable results.

Extreme Gradient Boosting (XGBoost)A total of 300 boosting estimators, a learning rate of 0.05, a maximum tree depth of 4, a subsampling ratio of 0.8, and a feature subsampling ratio of 0.8 make up the XGBoost model. Using the evaluation metric 'logloss' and fixing the random seed value, together with setting scale_pos_weight = spw, corrects the class imbalance.

4. Light Gradient Boosting Machine (LightGBM)

A total of 300 estimators, a learning rate of 0.05, and fifteen leaves per tree make up its implementation. To rectify the class imbalance, we set class_weight = 'balanced' and select a subsampling ratio of 0.8 for both features and subsampling. Turned off verbose output and fixed random seed.

5. Categorical Boosting (CatBoost)

The parameters for CatBoost are as follows: learning rate 0.05, depth of the tree 4, and 300 iterations. By setting auto_class_weights = 'Balanced', we may address the issue of class imbalance. The reproducible results and controllable output and file generation were achieved using the options random_seed = SEED, verbose = 0, and allow_writing_files = False.

G. Experimental Settings and Model Assessment

The experiments used a fixed random seed of 42 and a stratified 80:20 training–test split. PCA retained 90% cumulative explained variance. The candidate number of clusters ranged from 2 to 8, with final selection restricted to at least three clusters. Consensus clustering combined 15 base clusterings, including ten repeated K-Means runs, and stability was assessed using ARI over 80% subsamples. A minimum cluster-size threshold of 3% was applied, with K-Means labels used as a fallback. The Python and library versions should be documented based on the execution environment. A machine with a 3.6 GHz Intel W-2133 CPU, an A100 GPU, and 128 GB of RAM is used for the research. It is possible to assess the proposed ML models using metrics like Brier Score, ACC, PRE, REC, F1, Specificity, ROC - AUC, PRE-REC Area under the curve (PR-AUC), and Matthews correlation coefficient (MCC). ACC is the percentage of customers that are correctly classified, and PRE is the proportion of predicted positive reactions that really are positive. The ratio of correct samples to total positive samples is known as REC. The F1 is determined by averaging the PRE and REC values harmonically. The capacity to accurately identify clients who do not respond is known as specificity. Every one of the four confusion matrix results contribute to an evaluation of MCC's overall classification quality. Equation shows how this assessment measures are expressed mathematically (1-5).

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+Fp+TN+FN}} \quad (1)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (2)$$

$$Recall = \frac{\mathrm{TP}}{TP+FN} \quad (3)$$

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

$$MCC = \frac{(TP \times TN)-(FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

PR-AUC, which is particularly useful for unbalanced data, illustrates how the trade-off between PRE and REC varies over thresholds, whereas ROC-AUC measures the model's performance at all thresholds. The Brier Score is a way to evaluate how well the anticipated response probability match up with the observed ones; lower numbers indicate higher ACC.

4. Experimental Results

The marketing campaign's reactions were predicted using the Customer Personality Analysis dataset, and Table I displays the comparative performance of seven ML classifiers. Of the given metrics, GB stands out with an ACC of 0.8759, while LR ranks first in balanced ACC (0.8481), REC (0.8571), F1 (0.6243), MCC (0.5665), ROC-AUC (0.9090), and PR-AUC (0.6701). The GB model outperforms the others in terms of ACC in response prediction, as evidenced by its lowest Brier score of 0.0888.

Results from the seven ML models' assessments using 5-fold cross-validation on the training data are presented in Table II. The top model overall is GB with an ACC of 0.883, but CatBoost has higher F1 (0.620), ROC-AUC (0.913), and MCC (0.551) than the competition. In terms of percentage of REC (0.808), PR-AUC (0.685), and balanced ACC (0.823), LR stands out as the top performer when it comes to forecasting a campaign candidate.

TABLE I. COMPARATIVE PERFORMANCE EVALUATION OF MACHINE LEARNINGT MODELS FOR MARKETING CAMPAIGN RESPONSE PREDICTION

Model	Accuracy	Balanced Accuracy	Precision	Recall	Specificity	F1-Score	MCC	ROC-AUC	PR-AUC	Brier Score

Logistic Regression	0.8418	0.8481	0.4909	0.8571	0.8391	0.6243	0.5665	0.9090	0.6701	0.1221
Gradient Boosting	0.8759	0.6927	0.6429	0.4286	0.9569	0.5143	0.4585	0.8762	0.6208	0.0888
CatBoost	0.8589	0.7997	0.5294	0.7143	0.8851	0.6081	0.5331	0.8829	0.6046	0.1054
XGBoost	0.8467	0.7275	0.5000	0.5556	0.8994	0.5263	0.4360	0.8854	0.5730	0.1052
LightGBM	0.8491	0.7094	0.5079	0.5079	0.9109	0.5079	0.4189	0.8723	0.5425	0.1081
Extra Trees	0.8589	0.7022	0.5455	0.4762	0.9282	0.5085	0.4279	0.8779	0.5035	0.1010
Random Forest	0.8589	0.6437	0.5676	0.3333	0.9540	0.4200	0.3617	0.8697	0.4983	0.0976

TABLE II. FIVE-FOLD CROSS-VALIDATED PERFORMANCE OF MACHINE LEARNING MODELS

Model	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC_AUC**	**PR_AUC**	**Bal_Acc**	**MCC**
LogReg	0.833 +/- 0.035	0.478 +/- 0.063	0.808 +/- 0.055	0.598 +/- 0.057	0.908 +/- 0.025	0.685 +/- 0.078	0.823 +/- 0.037	0.531 +/- 0.070
CatBoost	0.871 +/- 0.021	0.574 +/- 0.077	0.688 +/- 0.064	0.620 +/- 0.038	0.913 +/- 0.019	0.681 +/- 0.089	0.796 +/- 0.026	0.551 +/- 0.048
XGBoost	0.878 +/- 0.021	0.605 +/- 0.076	0.612 +/- 0.056	0.605 +/- 0.049	0.903 +/- 0.018	0.661 +/- 0.091	0.769 +/- 0.028	0.535 +/- 0.059
LightGBM	0.878 +/- 0.025	0.621 +/- 0.110	0.580 +/- 0.046	0.594 +/- 0.052	0.904 +/- 0.017	0.658 +/- 0.075	0.756 +/- 0.025	0.527 +/- 0.070
GradBoost	0.883 +/- 0.020	0.685 +/- 0.109	0.428 +/- 0.059	0.527 +/- 0.076	0.907 +/- 0.018	0.657 +/- 0.086	0.696 +/- 0.036	0.480 +/- 0.092

ExtraTrees	0.879 +/- 0.018	0.630 +/- 0.076	0.492 +/- 0.098	0.549 +/- 0.081	0.885 +/- 0.011	0.610 +/- 0.064	0.720 +/- 0.049	0.488 +/- 0.088
RandomForest	0.873 +/- 0.022	0.678 +/- 0.158	0.304 +/- 0.081	0.420 +/- 0.108	0.886 +/- 0.017	0.605 +/- 0.084	0.639 +/- 0.046	0.397 +/- 0.126

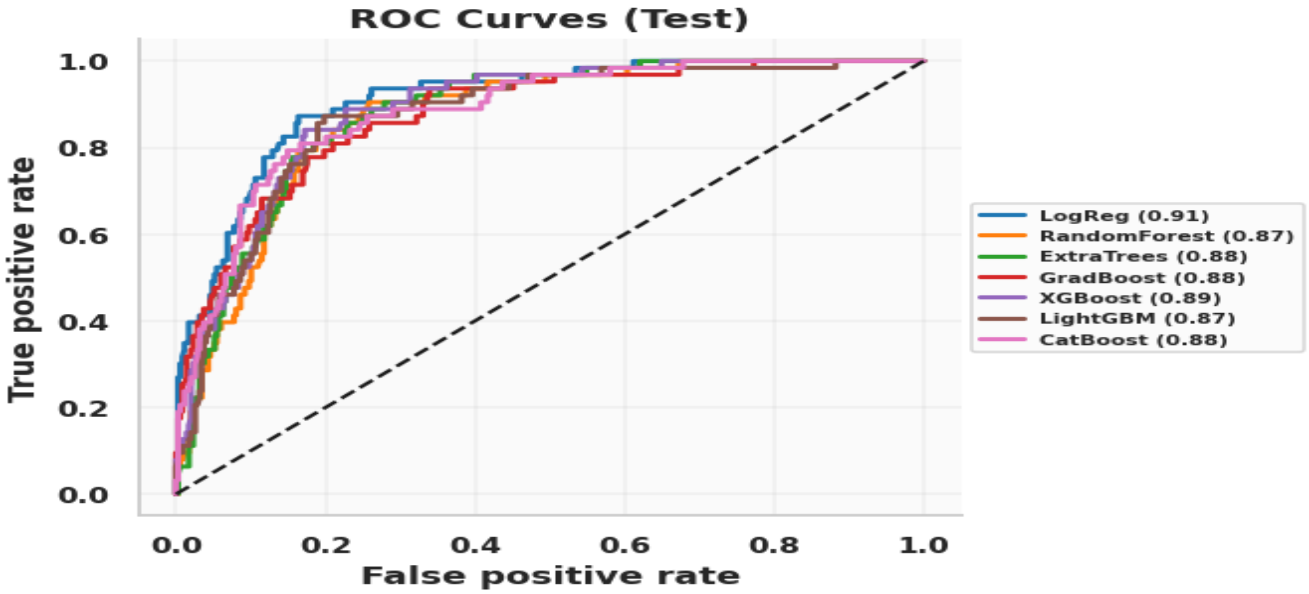


Fig. 8. ROC Curves of Machine Learning Models on the Test Set

Figure 8 shows the ROC curves of seven ML models that predicted the reaction to a marketing campaign. Different categorisation thresholds are illustrated by the curves, which show the trade-off between the TPR and the FPR. The top ROC-AUC score is 0.91 for LR, followed by 0.89 for XGBoost, Extra Trees, GB, CatBoost, RF, and LightGBM.

H. Profit-Based Decision Threshold Optimization and Campaign Performance Evaluation

The suggested methodology uses cross-validation prediction out-of-fold (OOF) to investigate the best Logistic Regression model threshold. The ideal profit, F1-optimal, and break-even points are also considered. Based on the data, the ideal profit barrier is 0.64, the ideal F1 threshold is 0.73, and the break-even point is 0.273. Classifying the test-set probability and determining campaign profitability are both affected by the selected profit threshold. In terms of profit per 100 customers, percentage of customers reached, customer categorisation metrics, and confusion matrix, the framework contrasts model-based targeting with contacting all consumers. It also analyses and compares the outcomes. The expected campaign profit is calculated as in equation 6:

$$Profit = N_t(rp_r - c) \quad (6)$$

where N_t denotes the number of contacted customers, r is the assumed net return per response, p_r is the predicted response probability, and c is the contact cost per customer. The break-even response probability is $p_{BE}=c/r=0.273$, indicating that contacting a customer is expected to break even when the response probability reaches 27.3%. The reported profit of 61.1 currency units per 100 customers represents an estimated outcome under these assumptions rather than an observed commercial return. The assumed numerical values of r and c must be stated explicitly to ensure reproducibility.

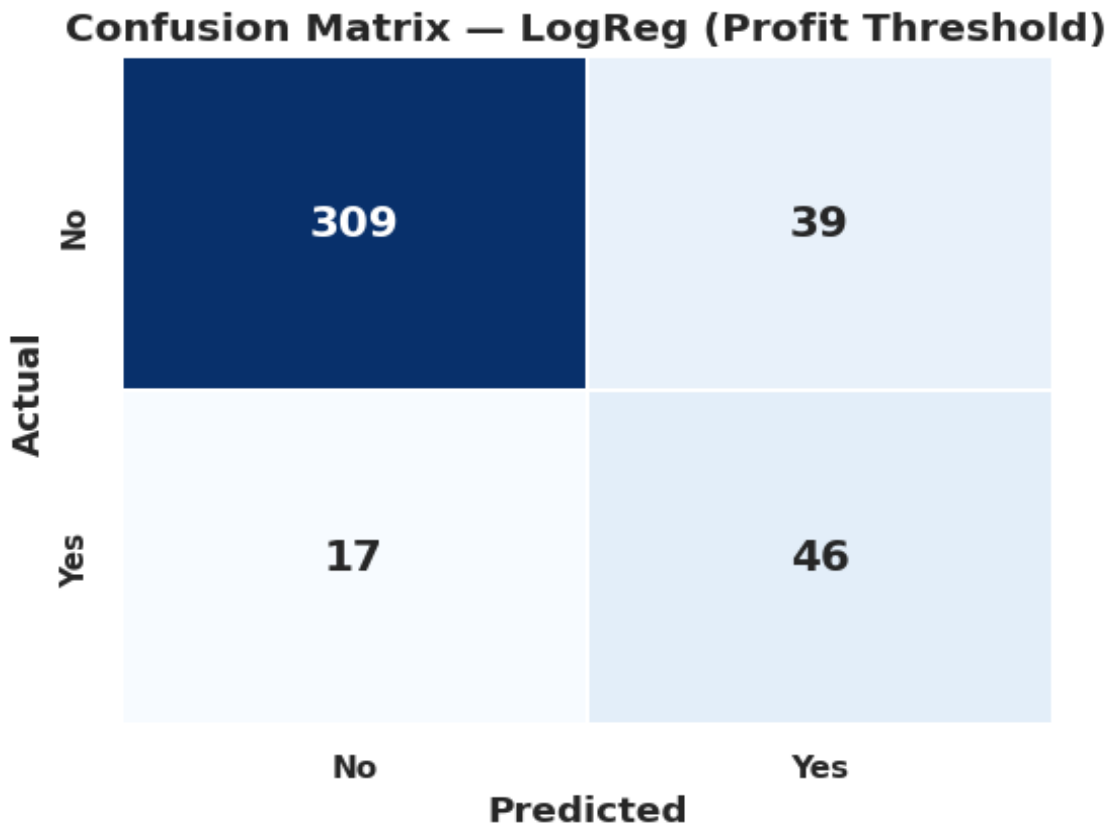


Fig. 9. Confusion Matrix of Logistic Regression

Figure 9 shows the LR model's confusion matrix that accounts for the optimal profit threshold. Though it erroneously identifies 39 non-responders as responding (FP) and 17 responders as non-responding (FN), the model effectively categorises 309 TN and 46 TP clients. Overall, the model's specificity rate is 88.79% and REC rate is 73.02%, demonstrating its ability to collect campaign responders with little needless customer encounters.

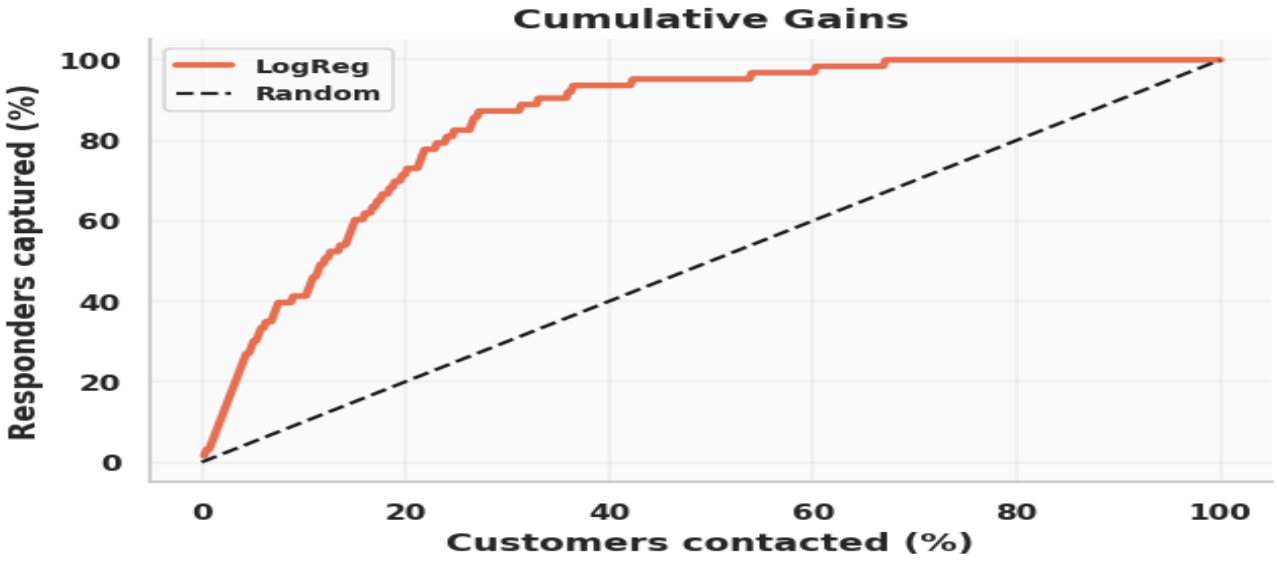


Fig. 10. Cumulative Gains Curve for Logistic Regression

The cumulative benefits of the LR model are higher than those of the randomly selected clients, as seen in Figure 10. Contacting 23% of clients yields around 80% campaign respondents, while contacting 32% yields over 90%. However, with random targeting, customers are specifically targeted according to their proportion of the whole customer base. Figure 11 compares the probability calibrations for predictions of responses to marketing campaigns using LR and RF. In a perfectly calibrated system, the dashed diagonal line represents the point where the anticipated and observed probabilities are identical. In terms of probabilistic ACC, RF outperforms LR, which has a score of 0.122.

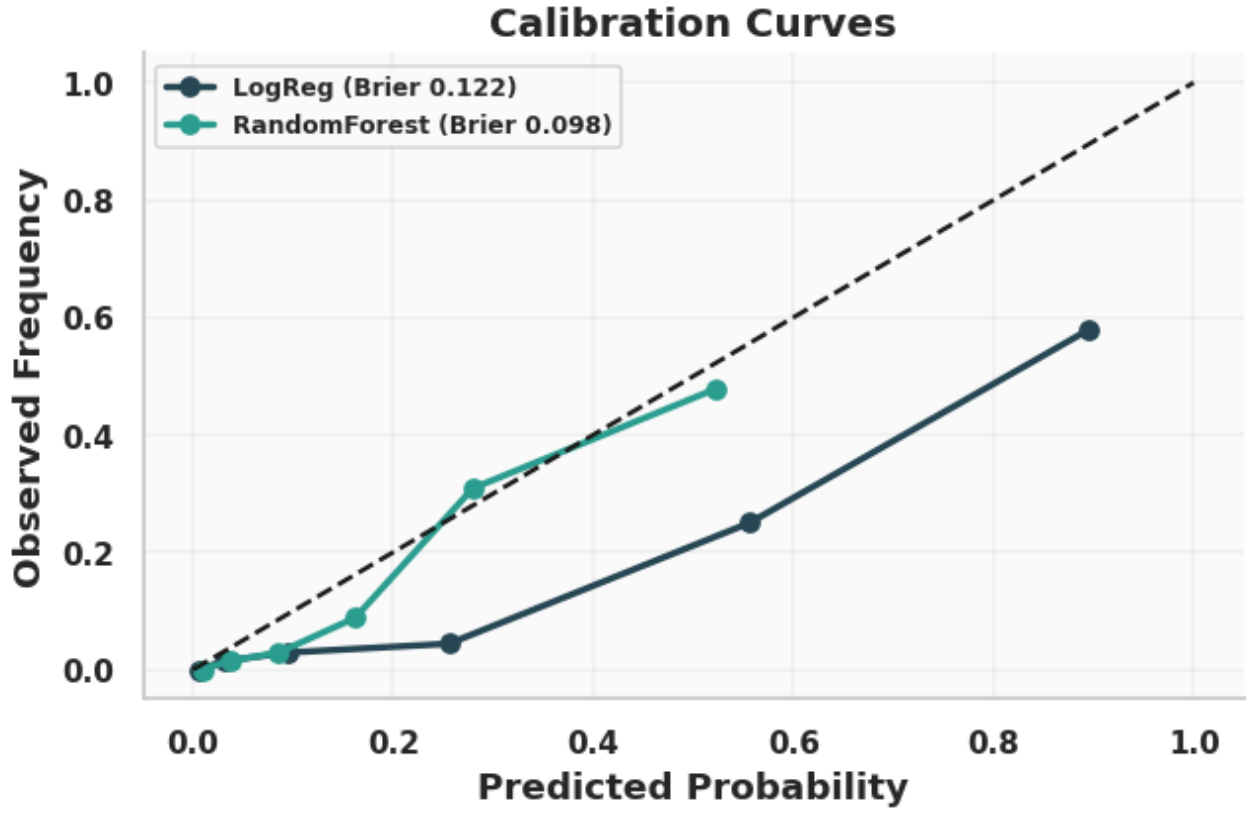


Fig. 11. Calibration Curves of Logistic Regression and Random Forest Models

Table III compares the baseline Logistic Regression model using 45 features with the segment-aware Logistic Regression model using 51 features on the held-out test set. Thresholds of 0.640 and 0.620, respectively, were selected using training-set out-of-fold predictions to maximize estimated profit. The segment-aware model improved recall from 0.730 to 0.794 and F1-score from 0.622 to 0.649, while PR-AUC increased from 0.670 to 0.684.

TABLE III. PERFORMANCE COMPARISON OF BASELINE AND SEGMENT-AWARE CAMPAIGN RESPONSE MODELS

Model	No. of Features	Threshold	Accuracy	Balanced Accuracy	Precision	Recall	Specificity	F1-Score	MCC	ROC-AUC	PR-AUC	Brier Score
Baseline Respon	45	0.640	0.864	0.809	0.541	0.730	0.888	0.622	0.550	0.909	0.670	0.122

se Model												
Segment-Aware Response Model	51	0.620	0.869	0.838	0.549	0.794	0.882	0.649	0.586	0.909	0.684	0.122

Figure 12 displays the local SHAP explanations for the client who is anticipated to respond to the marketing effort with a chance of 1.00. Family Size and NumCatalogPurchases are the next most important positive contributors, after AcceptedCmp2 (campaign approval previously). Tenure days, AcceptedCmp1, and AcceptedCmp5 are the other good criteria. But Premium Ratio and Recency both have negative predictive power. The blue bars show negative SHAP contributions, whereas the orange bars show positive ones.

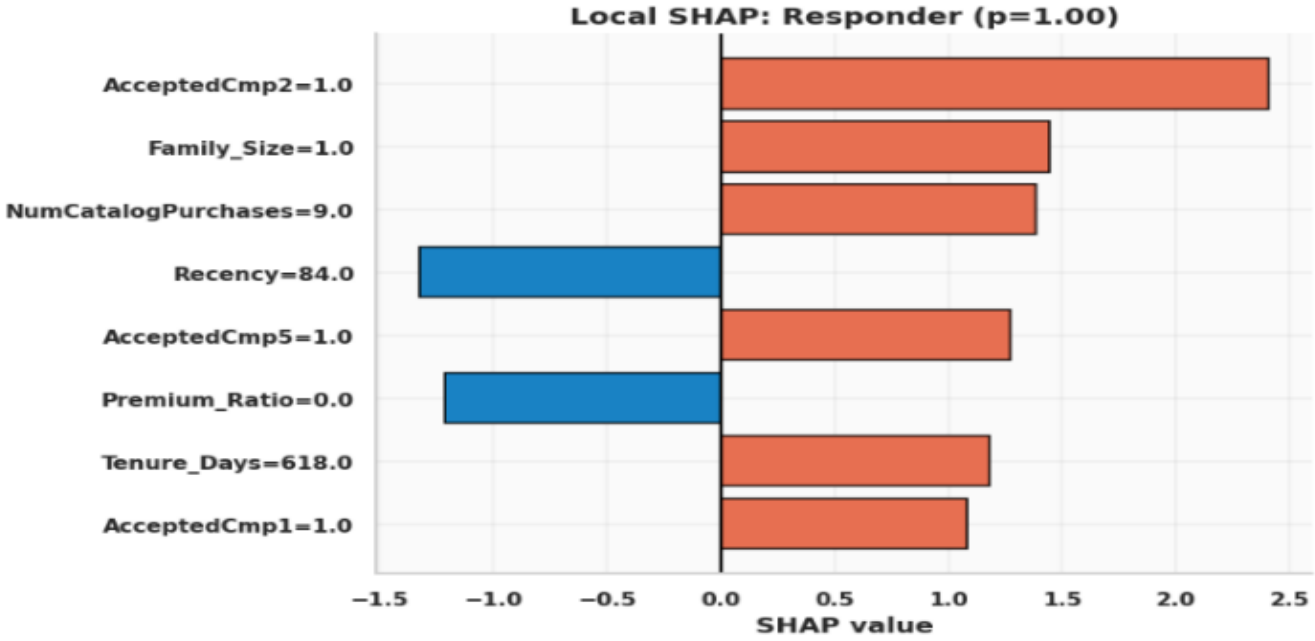


Fig. 12. Local SHAP Analysis of an Individual Customer's Campaign Response Prediction

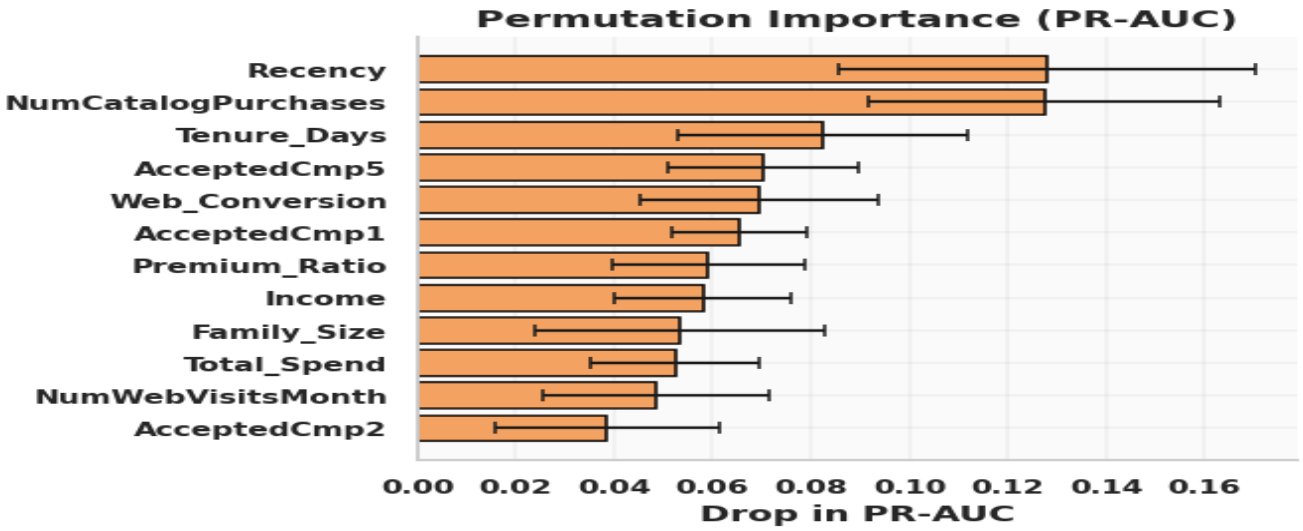


Fig. 13. Permutation Feature Importance Based on PR-AUC

The significance of customer characteristics is demonstrated in Figure 13 by examining the decline in PRE-REC-AUC following feature permutation. Ranking from most important to least important, with an approximate PR-AUC decrease of 0.128 for Recency and NumCatalogPurchases, we find Tenure Days at 0.082, AcceptedCmp5 at 0.070, and Web Conversion at 0.069. Income, AcceptedCmp2, NumWebVisitsMonth, Total_Spend, Family_Size, Premium_Ratio, and AcceptedCmp1 are other important characteristics.

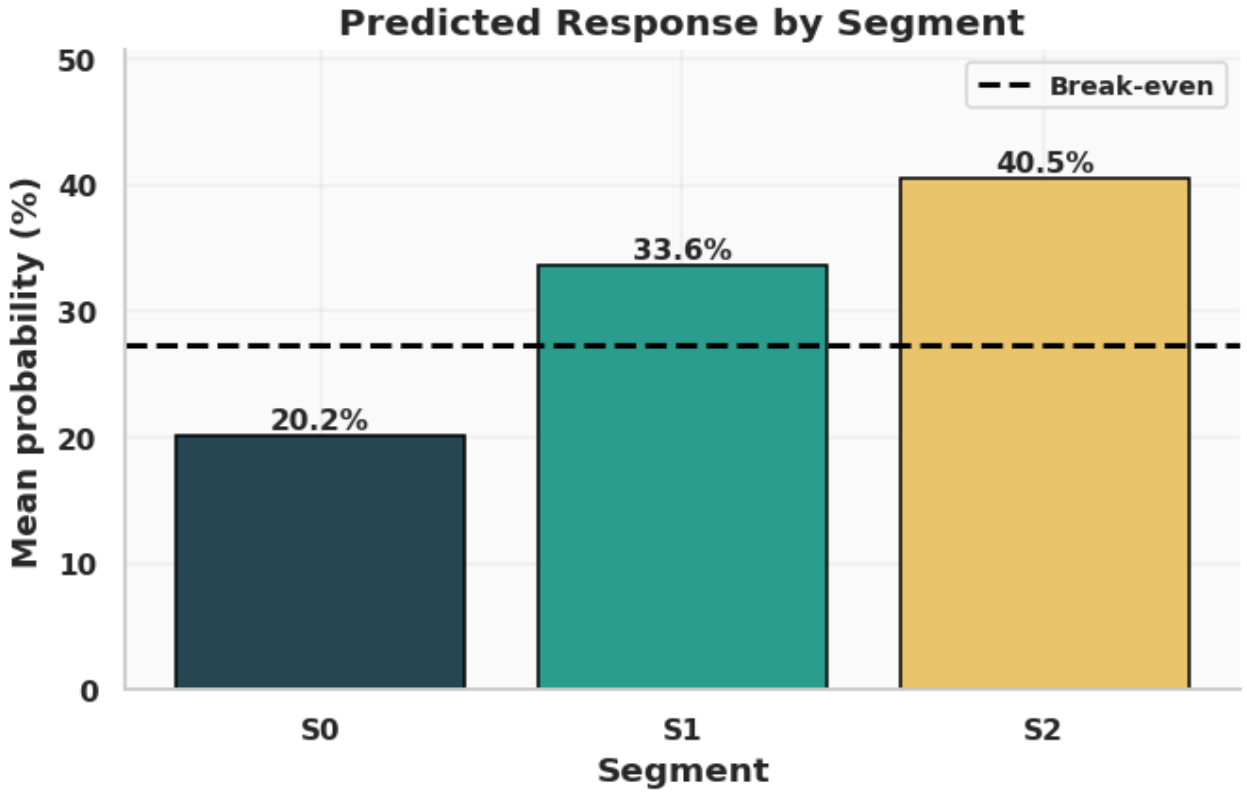


Fig. 14. Predicted Marketing Campaign Response Probability

Figure 14 shows a comparison of the three customer groups (S0, S1, and S2) with respect to the average expected campaign response probability. With Segment S0, the estimated response probability is the lowest at 20.2%, followed by S1 at 33.6%, and with Segment S2 at 40.5%. Attaining breakeven responses of around 27.3% is indicated by the dashed horizontal line. This barrier is exceeded in segments S1 and S2, which have a better possibility for lucrative targeting in campaigns, whereas S0 is below the threshold.

The mean absolute SHAP values of important factors impacting client segmentation are illustrated in Figure. 15. The most important characteristics for S0 are Total Spend and Total Purchases, which both contribute 0.12, whereas the most important feature for S2 is Deal Sensitivity, which contributes 0.12. S1 receives 0.08 from Total Purchases, 0.07 from Total Spend, and 0.07 from Deal Sensitivity. In addition to income and web conversion, numwebvisits per month also play a role in segmentation.

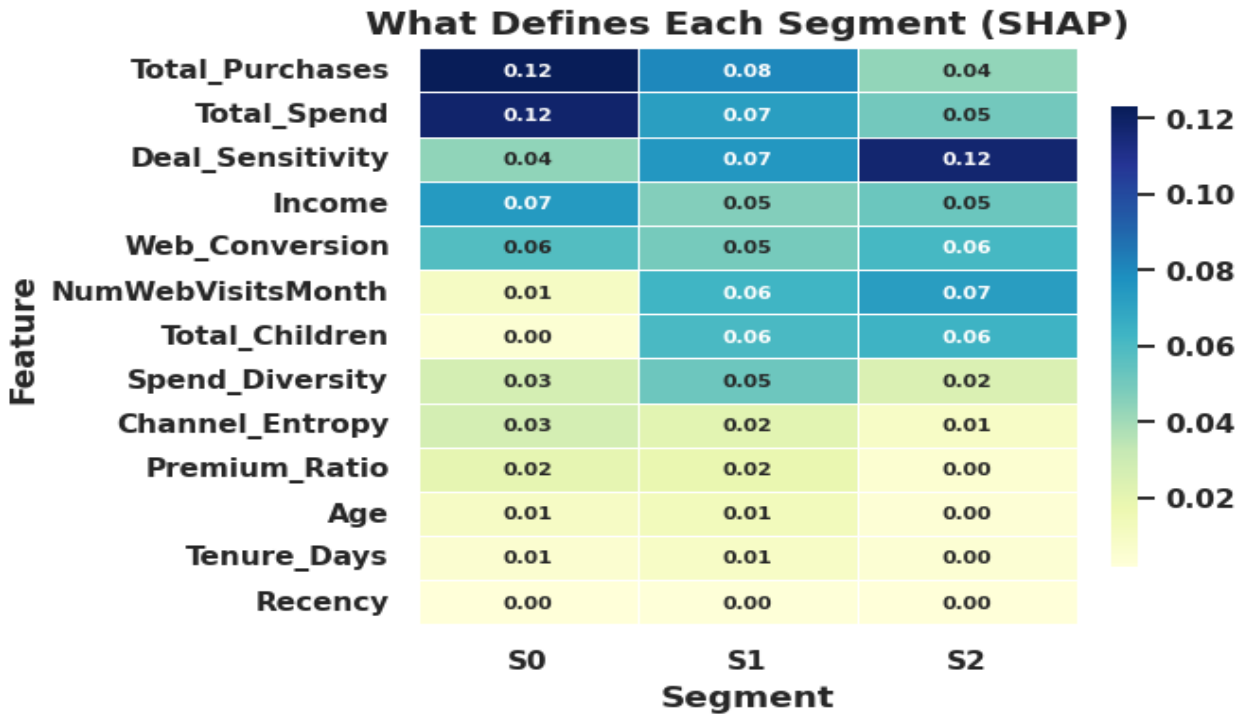


Fig. 15. SHAP-Based Feature Importance Across Customer Segments

Table IV shows the segment-level outcomes of the campaign's targeting and response for three different types of customers. The actual response rate for the segment S1 was the lowest (16.1%), the actual response rate in the segment S2 was the highest (24.6%) and the mean projected response probability in the segment S2 was the highest (40.5%). This can lead to the possibility of direct marketing campaigns reaching both segments. The goal of targeting is clear: it is best to target by individual model scores.

Table V shows the top 10 consumers, as predicted by the model using the test set, and their SHAP values indicate why they are the most likely to respond to the ad. The response probabilities are relatively high (0.980–0.998). Among the top ten customers, seven belong to S2, two belong to S1, and one customer (ID 7573) belongs to S0. The premium product ratio, recency and monthly visits to the website were all negative influencers for the individual forecasts, but the top ranked positive influencers were acceptance of previous campaigns (AcceptedCmp2 and AcceptedCmp5) and catalogue purchases.

TABLE IV.

TABLE V. SEGMENT-LEVEL CAMPAIGN RESPONSE AND CUSTOMER TARGETING SUMMARY

Customer Segment	Customers	Actual Response Rate	Mean Predicted Probability	Targeted Customers	Recommended Action
S0: Budget Families, Deal-seekers	**157**	**7.0%**	**20.2%**	**15**	**Selective targeting using model scores**
S1: Core Families, Full-price	**124**	**16.1%**	**33.6%**	**27**	**Prioritize for direct campaigns**
S2: Premium Child-free, Full-price	**130**	**24.6%**	**40.5%**	**43**	**Prioritize for direct campaigns**

TABLE VI. TOP 10 HIGH-PRIORITY CUSTOMERS FOR MARKETING CAMPAIGN TARGETING WITH SHAP-BASED EXPLANATIONS

Customer ID	Segment	Actual Response	Response Probability	Target Flag	Top Positive Driver	Positive SHAP	Top Negative Driver	Negative SHAP
3138	S2: Premium Child-free, Full-price	1	0.998	1	AcceptedCmp2	2.410	Recency	−1.327
10446	S2: Premium Child-free, Full-price	1	0.997	1	AcceptedCmp5	1.274	NumWebVisitsMonth	−1.170
10140	S2: Premium Child-free, Full-price	0	0.994	1	AcceptedCmp2	2.410	Premium_Ratio	−1.544
7627	S2: Premium Child-free, Full-price	1	0.994	1	NumCatalogPurchases	1.560	Premium_Ratio	−1.344

830	S2: Premium Child-free, Full-price	1	0.994	1	Recency	1.893	Tenure_Days	−1.182
7573	S0: Budget Families, Deal-seekers	1	0.992	1	Premium_Ratio	1.561	Web_Conversion	−0.619
7999	S2: Premium Child-free, Full-price	1	0.987	1	AcceptedCmp5	1.274	Premium_Ratio	−1.340
3584	S1: Core Families, Full-price	1	0.985	1	NumCatalogPurchases	1.476	Premium_Ratio	−0.996
3174	S1: Core Families, Full-price	1	0.983	1	AcceptedCmp2	2.410	Premium_Ratio	−1.200
5848	S2: Premium Child-free, Full-price	1	0.980	1	AcceptedCmp2	2.410	NumWebVisitsMonth	−1.170

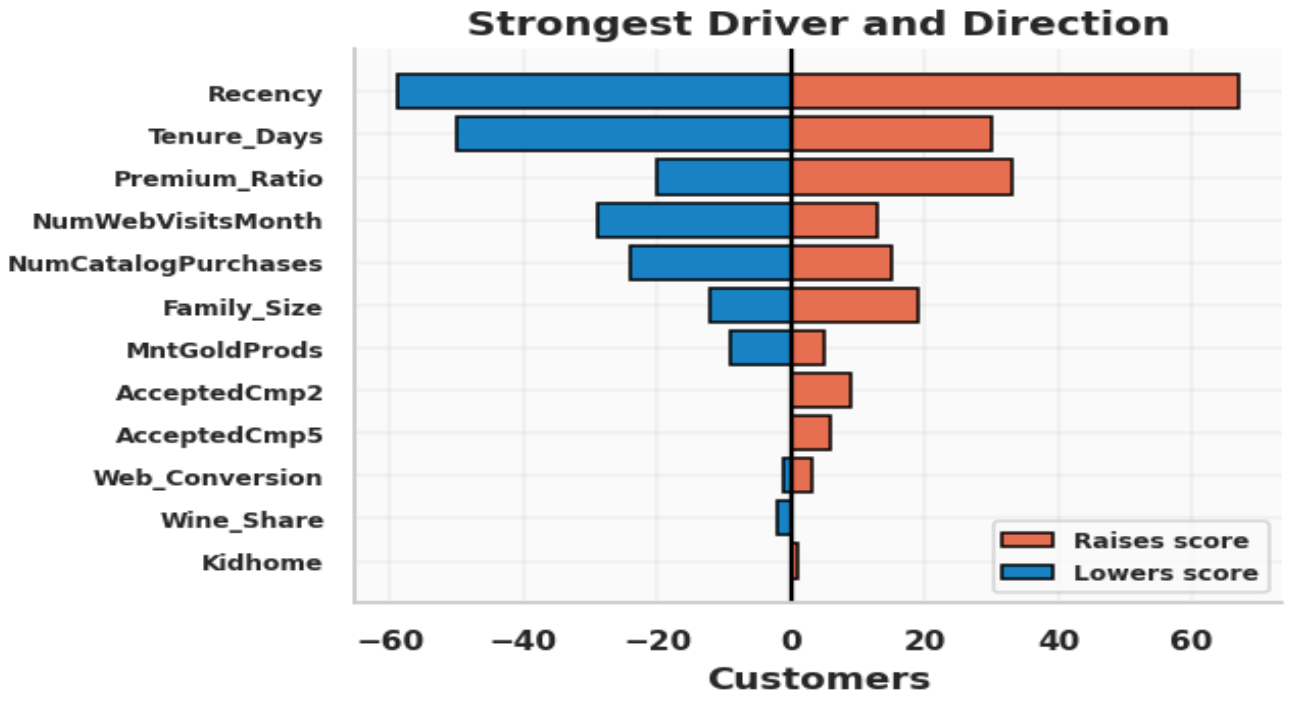


Fig. 16. Strongest Positive and Negative Drivers of Marketing Campaign

Figure 16 displays the highest and lowest consumer qualities that influence the expected campaign response scores in a good and negative way, respectively. About sixty-seven customers find recency to be the most helpful, while fifty customers find tenure days to be the most detrimental. When it comes to Premium Ratio, NumWebVisitsMonth, and NumCatalogPurchases, the other consequences are big and substantial. Previous campaign acceptance features (AcceptedCmp2 and AcceptedCmp5) had a larger impact on response scores than Family Size and MntGoldProds, two additional features from the same campaign. characteristics that raise the prediction score are depicted by the orange bars, whereas characteristics that lower the score are shown by the blue bars.

Figure 17 shows a comparison of seven ML models based on the predicted marketing profit per hundred clients. Considered in terms of profit, Logistic Regression earns 61.1 currency units, whereas CatBoost comes in second with 60.6. A profit of 53.8 is made by Extra Trees and 53.0 by LightGBM, whereas XGBoost's profit is 52.3. Gradient and RF Profits from increasing yield are somewhat modest at 49.4 and 47.0 percent, respectively.

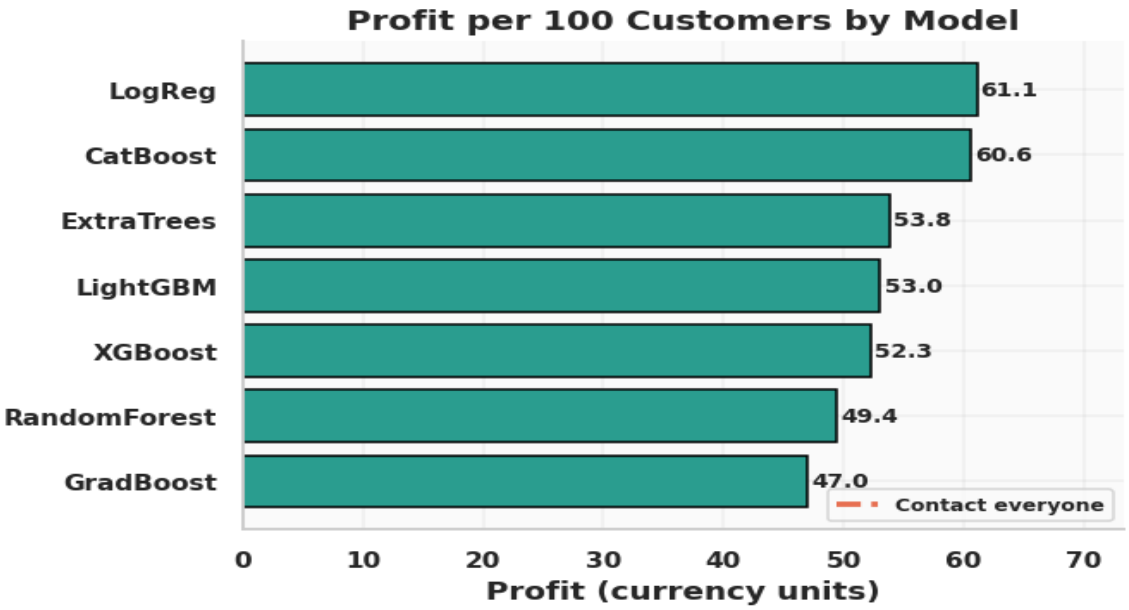


Fig. 17. Comparison of Profit per 100 Customers Across Machine Learning Models

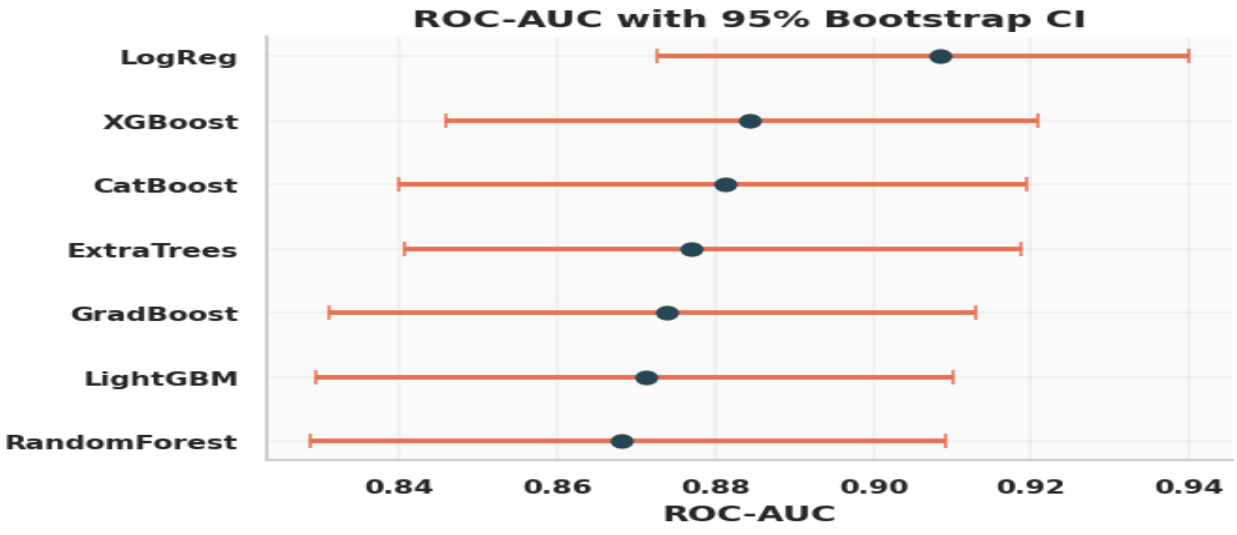


Fig. 18. ROC-AUC Comparison of Machine Learning Models with 95% Bootstrap

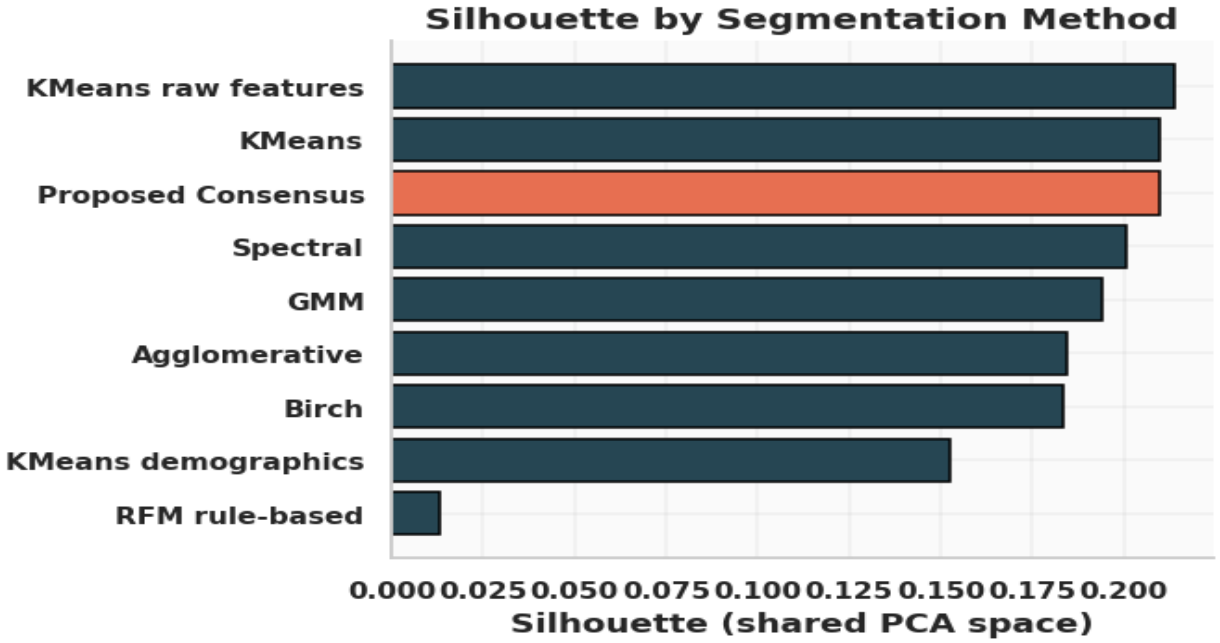


Fig. 19. Silhouette Score Comparison

Figure 18 shows the ROC-AUC of seven ML models, with 95% bootstrap confidence intervals indicated by horizontal error bars. Following Logistic Regression in terms of ROC-AUC score is CatBoost (0.883), XGBoost (0.886), Extra Trees (0.877), GB (0.874), LightGBM (0.871), and RF (0.868). At the top of the list is LR. Estimates of discrimination performance at the model level are accompanied with confidence intervals that show the extent of uncertainty in those estimates.

Figure 19 shows a comparison of the nine customer segmentation procedures utilising silhouette scores in the common PCA space. The suggested consensus technique

yields a score of about 0.210, while the K-Means algorithm (raw features) achieves the greatest score of about 0.215. Not only that, but RFM rule-based segmentation has the worst performance, whereas Spectral clustering and Gaussian Mixture Models both have competitive results.

Table VI shows the conclusions drawn from the Logistic Regression model's segment-wise Predictive Performance & Profitability on the Test set. The second segment (S2) has the best actual response rate, mean predicted probability, REC, and F1, and it also has the largest profit (129.23 currency units per 100 consumers). Segment S0's profit is 13.38 per 100 consumers, whereas Segment S1's profit is 50.00.

TABLE VII. SEGMENT-WISE PREDICTIVE PERFORMANCE AND PROFITABILITY OF LOGISTIC REGRESSION ON THE TEST SET

Segment	Customers	Actual_Resp_%	Mean_Pred_%	ROC_AUC	Precision	Recall	F1	Contacted_%	Model_Profit_per_100	ContactAll_Profit_per_100
S0	157	7.01	20.17	0.94	0.40	0.55	0.46	9.55	13.38	-222.93
S1	124	16.13	33.62	0.83	0.48	0.65	0.55	21.77	50.00	-122.58

S2	130	24.62	40.51	0.92	0.63	0.84	0.72	33.08	129.23	-29.23

The Logistic Regression model was evaluated through an ablation study using five-fold cross-validation on the training split, as shown in Table VII. The baseline model achieved a ROC-AUC of 0.908 ± 0.025 and a PR-AUC of 0.685 ± 0.078. Adding segment indicators or centroid distances did not improve PR-AUC, and combining both reduced it to 0.672. Removing campaign history reduced PR-AUC to 0.565, highlighting the predictive relevance of previous campaign interactions. Segment features alone achieved a ROC-AUC of 0.659, while removing value and RFM features reduced ROC-AUC to 0.861. These results indicate that segmentation features do not consistently improve overall discrimination in cross-validation, although they may improve responder identification at selected decision thresholds on the held-out test set.

TABLE VIII. ABLATION STUDY WITH TUNED BEST MODEL USING FIVE-FOLD CROSS-VALIDATION

Feature Set	**No. of Features**	**ROC-AUC (Mean ± SD)**	**PR-AUC (Mean ± SD)**	**F1-Score**	**Δ ROC-AUC**	**Δ PR-AUC**
Baseline (All Base Features)	45	0.908 ± 0.025	0.685 ± 0.078	0.598	0.000	0.000
Baseline + Segment One-Hot	48	0.908 ± 0.025	0.677 ± 0.086	0.600	+0.001	−0.008
Baseline + Centroid Distances	48	0.907 ± 0.024	0.679 ± 0.078	0.595	−0.000	−0.007
Segment-Aware (Both)	51	0.908 ± 0.025	0.672 ± 0.086	0.595	−0.000	−0.013
Only Segment Features	6	0.659 ± 0.044	0.262 ± 0.028	0.325	−0.249	−0.423
Only Campaign History	7	0.718 ± 0.016	0.399 ± 0.028	0.460	−0.190	−0.286
Without Campaign History	38	0.874 ± 0.034	0.565 ± 0.067	0.541	−0.034	−0.121
Without Demographics	32	0.882 ± 0.018	0.620 ± 0.059	0.526	−0.026	−0.065
Without Value and RFM	38	0.861 ± 0.029	0.575 ± 0.071	0.492	−0.046	−0.111

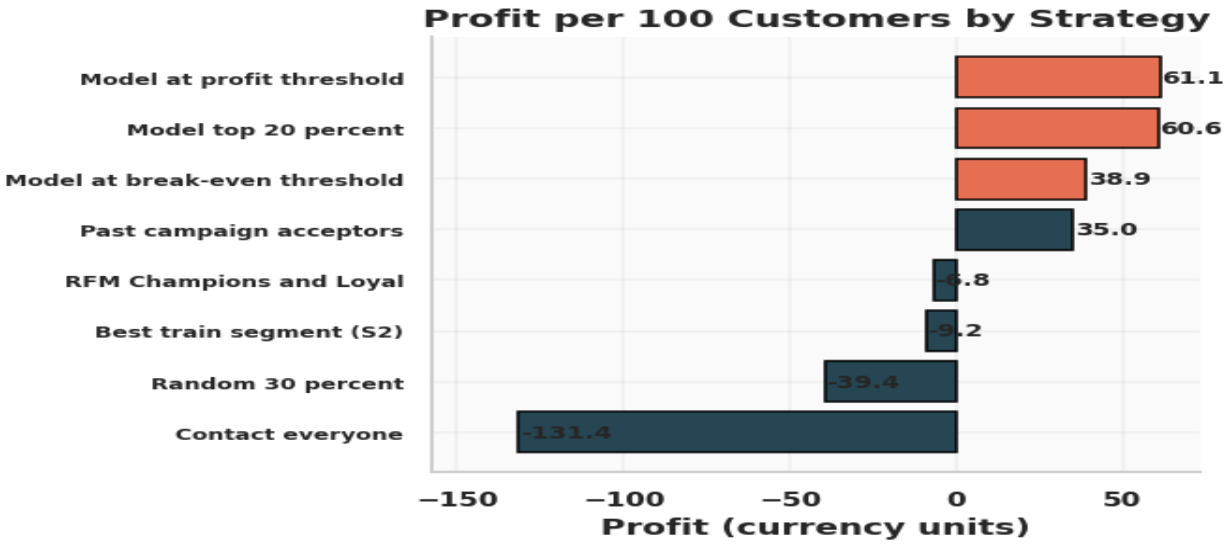


Fig. 20. Profit Comparison of Customer Targeting Strategies

Figure 20 shows a comparison of eight different marketing methods based on the profit gained per hundred consumers. At 61.1 currency units, the model-based profit threshold is the highest, with 60.6 being the second-highest, achieved by focusing on the top 20% of consumers. After accounting for prior campaign acceptors, the break-even point is 38.9, and 35.0 is the result of focusing on them. As a comparison, contacting all consumers leads to the biggest loss of 131.4, while randomly selecting 30% of customers results in a loss of 39.5.

The results presented in Table VIII should be interpreted as contextual comparisons rather than direct performance benchmarks. The cited studies use different datasets, sample sizes, feature sets, class distributions, preprocessing procedures, and evaluation protocols. Therefore, their reported performance metrics are not directly comparable with the results obtained in this study. LR achieved a ROC-AUC of 0.9090, an accuracy of 84.18%, and a REC of 85.71%. Although the REC is lower at 42.86%, GB obtains the greatest ACC at 87.59% and PRE at 64.29%. With an F1 of 0.731, Classifier F outperforms the NNs (79.00%) and the LAD Tree Algorithm (76.08%).

TABLE IX. COMPARATIVE PERFORMANCE OF EXISTING CLASSIFICATION MODEL

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.8418	0.4909	0.8571	0.6243	0.9090
Gradient Boosting	0.8759	0.6429	0.4286	0.5143	0.8762
Decision Tree[13]	-	-	-	-	0.842
Classifier f[14]	0.808	-	-	0.731	-
NN[15]	0. 79	-	-	-	-

LAD tree algorithm[16]	0.7608	-	-	-	-

The study found that ML algorithms can effectively predict customer preferences, buying patterns, and consumer behaviour, which enables more precise customer segmentation and a better prediction of campaign response. The greatest benefits to this are targeted marketing, avoiding overspending a campaign, and making better data driven decisions. The adoption of explainable AI tools improves transparency, and can also help determine what factors influence consumer responses. Moreover, companies can direct more of their advertising dollars towards the most likely audience for their ads, which can then be more targeted and customised.

B. Justification and Novelty

Traditional marketing methods targeting broad audiences are oblivious to customer behavior, intent to react, profitability of the campaign, and interpretability. This study overcomes these challenges by combining seven machine learning classifiers, optimizing the decision threshold according to the profit, and a single architecture for explainable artificial intelligence based on consensus-based consumer segmenting. The unique feature is the integration of behavioural segments into a single workflow to predict client actions, explain in terms of SHAP, and assess profitability of segments. The framework includes consideration of the value of segment information and the ranking of potential responders as a way to support individual marketing strategies, while placing emphasis on key reaction criteria. The results of the experiments demonstrate the superiority of the segment-aware model over the baseline model in terms of REC and F1 metrics, and the fact that campaign resources could be used more efficiently using the profit-based targeting method.

5. Conclusion

The suggested framework, which is built on Explainable AI, integrates the ideas of customer segmentation, campaign response prediction, and profit-driven targeting to support marketing decision-making. The LR model

achieved the highest ROC-AUC value of 0.909, and the highest test ACC value of 87.59% was obtained with the GB model. Profit based targeting achieved a 61.1 currency unit return for every 100 clients, suggesting that more ads that do not help the business can be decreased. In line with the objective to develop unique and tailored marketing strategies, SHAP and permutation significance revealed that campaign history, recency and catalogue purchase were important predictors. There are some limitations to the study's findings, however—such as limiting the data to a single Kaggle database, the class imbalance, and a lack of evidence to validate the findings for other populations or marketing contexts. A second is that explainable AI can discover connections, but not cause and effect, and that campaign profitability is dependent on income prediction from ‘response' minus campaign costs. Future research will assess the effectiveness of probability calibration, develop more sophisticated methods for dealing with imbalances and ensemble learning, and test the framework with larger and more real-world multi-industry datasets. Further research on adaptive campaign thresholds, real-time consumer behaviour and causal analysis could help enhance the marketing relevance, interpretability and robustness of models.

References

1. F. M. Talaat, A. Aljadani, B. Alharthi, M. A. Farsi, M. Badawy, and M. Elhosseini, “A Mathematical Model for Customer Segmentation Leveraging Deep Learning, Explainable AI, and RFM Analysis in Targeted Marketing,” *Mathematics*, vol. 11, no. 18, p. 3930, Sep. 2023, doi: 10.3390/math11183930.
2. P. T. M. Visakai and S. Venkatraghavan, “AI-Driven Customer Segmentation and Loyalty Prediction for Personalised Marketing: An Interpretable Machine-Learning Analysis of a Small, High-Collinearity Customer Dataset,” *Iconic Res. Eng. Journals*, vol. 10, no. 2, pp. 3446–3460, Sep. 2026, doi: 10.64388/IREV10I2-1722663.
3. J. Joung and H. Kim, “Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews,” *Int. J. Inf. Manage.*, vol. 70, p. 102641, Jun. 2023, doi: 10.1016/j.ijinfomgt.2023.102641.
4. M. Asfi, B. Warsito, and A. Wibowo, “COSHAP: An Explainable AI Framework Integrating Conformal Prediction and SHAP for Uncertainty-Aware Consumer Segmentation,” *J. Curr. Sci. Technol.*, vol. 16, no. 4, p. 213, Sep. 2026, doi: 10.59796/jcst.V16N4.2026.213.

5. S. Trivedi, M. Ramarajan, P. Sivaganga, M. Kumar, S. Sujatha, and S. Sangeetha, "An AI-Driven Predictive Framework for Marketing Strategy Optimization Using Consumer Behavior Analytics," in *2026 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, IEEE, Aug. 2026, pp. 1–6. doi: 10.1109/ASSIC70037.2026.11717336.

6. A. A. Durib and A. Gökmen, "An Explainable End-to-End AI Framework for Marketing Automation and Lead Generation in Small and Medium Enterprises," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 18, no. 4, pp. 450–467, 2026, doi: https://doi.org/10.70917/ijcisim-2026-2529.

7. S.Sowndharya and R.Hariharan, "Effect of AI-Driven Hyper Personalization vs Traditional Segmentation on Customer Engagement and Conversion in Digital Marketing," in *2026 6th International Conference on Inventive Computation and Information Technologies (ICICIT)*, IEEE, Jul. 2026, pp. 1124–1129. doi: 10.1109/ICICIT69063.2026.11633971.

8. A. Khowaja, A. Sarfaraz, R. Zafar, M. Saqib, M. Ali, and S. F. F. Rizvi, "Customer Segmentation with Personalized Offers Using Machine Learning & If-Else Statements," in *2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC)*, IEEE, Apr. 2025, pp. 1–6. doi: 10.1109/ICETECC65365.2025.11071180.

9. M. El-Hajj and M. Pavlova, "Predictive Modeling of Customer Response to Marketing Campaigns," *Electronics*, vol. 13, no. 19, p. 3953, Oct. 2024, doi: 10.3390/electronics13193953.

10. S. Agarwal, S. Singh, B. Hazela, and V. Singh, "Harnessing Machine Learning for Effective Customer Segmentation," in *2024 International Conference on Signal Processing and Advance Research in Computing (SPARC)*, IEEE, Sep. 2024, pp. 1–6. doi: 10.1109/SPARC61891.2024.10829373.

11. N. Godcares, A. Sirsath, A. Bongale, P. Kadam, R. Jayawal, and S. Patil, "Exploring Customer Segmentation in the Context of Market Analysis," in *2023 IEEE 11th Region 10 Humanitarian Technology Conference (R10-HTC)*, IEEE, Oct. 2023, pp. 444–449. doi: 10.1109/R10-HTC57504.2023.10461815.

12. A. Patel, "Customer Personality Analysis," Kaggle. [Online]. Available: https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

13. V. Mishra, "BUSINESS ANALYTICS FOR CUSTOMER SEGMENTATION AND MARKETING CAMPAIGN EFFECTIVENESS : A MACHINE LEARNING APPROACH," *IJRDO-Journal Bus. Manag.*, vol. 12, no. 02, 2026, doi: 10.69980/bm.v12i2.6768.

14. S. K. Jauhar, S. Harinath, V. Krishnaswamy, and S. K. Paul, "Explainable artificial intelligence to improve the resilience of perishable product supply chains by leveraging customer characteristics," *Ann. Oper. Res.*, vol. 354, no. 1, pp. 103–142, Nov. 2025, doi: 10.1007/s10479-024-06348-z.

15. S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decis. Support Syst.*, vol. 62, pp. 22–31, Jun. 2014, doi: 10.1016/j.dss.2014.03.001.

16. A. G. Rekha, M. S. Abdulla, and S. Asharaf, "Artificial Intelligence Marketing: An application of a novel Lightly Trained Support Vector Data Description," *J. Inf. Optim. Sci.*, vol. 37, no. 5, pp. 681–691, Sep. 2016, doi: 10.1080/02522667.2016.1191186.

.