

Computational Modeling of Neural Networks for Predicting Consumer Behavior in Digital Marketing, Application in human resource management for training sales staff

Amirhossein Golzarizamir

Bachelor of Business Administration, Hamedan Branch, Islamic Azad University, Hamedan, Iran
amirhosseinglzar@gmail.com

Abstract

This study aims to design and evaluate a computational model based on multilayer perceptron (MLP) neural networks for predicting consumer purchasing behavior in digital marketing. Data from 50,000 users on a digital marketing platform were collected, including demographic variables (age, gender, location, device type), behavioral variables (click count, time spent, return rate, purchase history), and interactional variables (engagement with ads and content). The MLP model was designed with three hidden layers (64, 32, and 16 neurons) and ReLU activation function, and trained using the Adam optimization algorithm with a learning rate of 0.001. The dataset was split into 70% training, 15% validation, and 15% testing, with Dropout and batch normalization applied to mitigate overfitting. For interpretability, SHAP analysis and feature importance assessment were employed. The findings indicated that the proposed model achieved excellent predictive performance (Accuracy = 0.931, F1 = 0.922, Precision = 0.918, Recall = 0.926, ROC-AUC = 0.974). The performance difference between training and test sets was less than 1.5%, demonstrating high model generalizability. SHAP analysis revealed that three features—“purchase history” (24.6%), “number of ad clicks” (18.9%), and “time spent on site” (15.7%)—contributed most significantly to the prediction of purchase likelihood. Moreover, applying the model in simulated scenarios resulted in an average increase in advertising return on investment (ROI) of up to 15% and a reduction of 12.3% in costs due to inappropriate targeting. These results suggest that the designed neural network model can accurately identify complex purchasing behavior patterns and provide valuable managerial insights for data-driven decision-making in digital marketing.

Keywords: Artificial neural networks, deep learning, digital marketing, consumer behavior prediction, SHAP analysis, return on investment.

1. Introduction

The rapid pace of digital transformation in recent decades has not only reshaped economic activities and social interactions but has also fundamentally altered traditional business structures and marketing approaches. The digitization of processes, the rising penetration of the Internet, the widespread use of smartphones, and the emergence of technologies such as artificial intelligence, the Internet of Things, and big data have created a new landscape for interactions between brands and consumers [1]. This landscape is characterized by unprecedented levels of information exchange, content accessibility, and consumer choice [1–2]. Consequently, traditional marketing tools, logics, and strategies—primarily based on one-way media and mass communication—are no longer capable of adequately responding to the needs and

behaviors of today's consumers [3]. In this environment, digital marketing has emerged as a novel, technology-driven approach and has gained a prominent position within business strategies. Digital marketing encompasses a set of marketing activities conducted through digital channels, tools, and platforms—such as search engines, social networks, websites, mobile applications, and e-commerce platforms—to attract, persuade, engage, and retain customers [4]. A key feature of this form of marketing lies in its measurability, precise targeting, and ability to personalize messages, which distinguishes it from traditional marketing and enhances its effectiveness [5].

Within this dynamic context, consumer behavior has undergone significant transformation as well. Today, consumers encounter a vast volume of digital stimuli throughout their purchasing journey, including targeted advertisements, personalized recommendations, user-generated content, online reviews, and intelligent recommender systems [6]. The purchasing decision is no longer a linear or straightforward process; rather, it is the outcome of a complex interplay among psychological factors (motivation, attitudes, perception, loyalty), social factors (social feedback, influencers, virtual communities), and technology-driven factors (user experience design, recommendation algorithms, marketing automation) [7]. This complexity has rendered classical consumer behavior models—such as AIDA and AISAS—designed for traditional marketing environments insufficient for accurately explaining consumer decision-making dynamics in the digital sphere. The multiplicity of brand–consumer touchpoints, the prominent role of interactive data, and the continuous variability of online user behavior have intensified the need for more flexible, data-driven models [4].

Artificial Neural Networks (ANNs), inspired by the structure and functioning of the human brain, are capable of approximating nonlinear and multidimensional relationships among variables and extracting complex patterns from large-scale datasets [8]. In recent years, the development of deep learning models and optimization algorithms such as Adam and RMSProp has significantly enhanced the applicability of these networks in analyzing consumer behavior [9]. Numerous studies have demonstrated that neural network–based models can deliver substantially higher accuracy in predicting purchase intention and conversion rates—compared to classical approaches such as logistic regression, decision trees, or support vector machines [10–14]. Forecasting consumer behavior in digital marketing is of strategic importance not only from a theoretical perspective but also in terms of practical implications. In an environment where competition among brands to attract and retain users has intensified, the ability to predict consumers' purchase probability enables the optimal allocation of advertising budgets, the design of targeted marketing campaigns, and improvements in return on investment (ROI) [15]. Furthermore, the massive volume of data generated through user interactions on platforms such as online retail websites, social networks, and search engines provides an unprecedented opportunity for leveraging machine learning techniques and predictive modeling [14]. These datasets include variables such as the number of clicks, time spent on webpages, bounce rate, purchase history, and device type, which, combined with demographic attributes, form a multidimensional framework for understanding customer behavior [16].

A review of recent studies indicates that the evolution of applying neural networks in digital marketing can be described in three stages. The first stage is the adaptive phase, during which simple neural networks were compared with traditional statistical models [17]. The second stage, the deepening phase, focused on leveraging these models for sentiment analysis, targeted advertising, and user behavior analytics in social media environments [19]. The third stage is the predictive intelligence phase, in which more advanced deep learning architectures were integrated with interpretability techniques—such as SHAP and

LIME—to understand model decisions and extract managerial insights [20]. This trajectory reflects a shift from mere predictive modeling toward explanatory and decision-support analytics [21]. Despite these advancements, several challenges remain, including computational complexity, the risk of overfitting, the need for extensive datasets, and the issue of model “black-box” opacity [22]. Recent research has sought to address these challenges by employing techniques such as regularization, batch normalization, and sensitivity analysis to enhance numerical stability and model interpretability [23]. Consequently, the application of neural networks in digital marketing has progressed beyond the experimental stage and has become a practical tool for data-driven decision-making within organizations [21]. Accordingly, the present study aims to develop a computational framework based on a multilayer perceptron (MLP) feedforward neural network for predicting consumer purchase behavior in digital marketing contexts. The novelty of this research lies in integrating the computational accuracy of neural network models with interpretability features through SHAP analysis and feature-importance assessment.

2. Mathematical Modeling and Methodology

The proposed framework of this study is based on the mathematical and computational modeling of users’ purchasing behavior in the context of digital marketing. For this purpose, a Multilayer Perceptron (MLP) feedforward neural network was used to extract nonlinear relationships and hidden interactions among users’ behavioral, demographic, and engagement variables. In this research, the data were collected from user interactions with the Digital platform. These data included user activity logs, page visits, ad clicks, session duration, purchase history, and device type during the first quarter of 2024 (January to March). In total, 50,000 unique user samples with diverse purchasing behaviors were analyzed, of which approximately 22% resulted in a final purchase (Conversion). All steps—from data preparation to defining the network architecture, cost function, optimization algorithm, and evaluating model performance—were explained within a formal and reproducible framework.

2.1. Conceptual and Data Framework of the Model

The dataset used in this study is defined as follows:

$$D = \{(X_i, y_i)\}_{i=1}^m, X_i \in \mathbb{R}^n, y_i \in \{0,1\},$$

where X_i is the feature vector of the i -th user and y_i is the binary dependent variable indicating the purchase status (1 = purchase, 0 = no purchase). The input features (X_i) consist of three main categories:

- Demographic Features: age, gender, device type (mobile, desktop, tablet), and geographical location.
- Behavioral Features: number of ad clicks, time spent on product pages, user return rate, and previous purchase history.
- Engagement Features: level of user activity in advertising campaigns, social media engagement rate, and the number of times items were added to the shopping cart.

To standardize the scales, numerical features were normalized to the range [0,1] using Min–Max normalization:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)}.$$

Categorical features such as device type and traffic source (Organic / Referral / Paid) were represented using One-Hot Encoding. Since the ratio of purchase to non-purchase samples was imbalanced (approximately 1 to 4), to avoid model bias, the training weights of each class were calculated based on the following equation and incorporated into the cost function:

$$w_c = \frac{m}{2m_c}, c \in \{0,1\},$$

where m_c is the number of samples in class c and m is the total number of samples.

Before starting the learning process, the dataset was cleaned for missing data, anomalies, and outliers:

- Missing data were replaced using the K-Nearest Neighbors Imputation method ($k = 5$).
- Outliers were identified using the Interquartile Range (IQR) criterion and adjusted if deviating more than 3σ .
- Variables with correlations greater than 0.9 with other features (Multicollinearity) were removed to increase the numerical stability of the model.

The final dataset after preprocessing included 47,320 valid records and 10 final independent variables. The data were split using Stratified Sampling into 70% for training, 15% for validation, and 15% for testing to preserve the class distribution in each subset.

The training and evaluation process of the model was implemented in the Python 3.10 environment using the PyTorch, Scikit-learn, and Panda's libraries. All experiments were conducted on a system equipped with an NVIDIA RTX 3080 GPU, a 12-core processor, and 64GB of RAM. The average training time per epoch was approximately 0.48 seconds, and the total training time for 250 epochs was around 120 seconds. To ensure computational stability, all experiments were repeated with a fixed random seed (seed = 42) and the average of five independent runs was reported.

This dataset and computational settings provided the basis for developing the mathematical model in the next section and evaluating its performance in the empirical analysis.

2.2. Mathematical Structure of the Neural Network Model

The proposed model was designed and implemented based on a Multilayer Perceptron (MLP) feedforward neural network. The final network consisted of three hidden layers with dimensions [64,32,16] neurons, and a single-neuron output layer predicting the users' purchase probability. The forward propagation process for each input sample X_i was defined as follows:

$$a^{(0)} = X_i, a^{(l)} = f(W^{(l)}a^{(l-1)} + b^{(l)}), l = 1,2,3,$$

where $W^{(l)} \in \mathbb{R}^{n_l \times n_{l-1}}$ denotes the weight matrix, $b^{(l)} \in \mathbb{R}^{n_l}$ the bias vector, and $f(\cdot)$ the activation function. In this model, the ReLU activation function was used in all hidden layers:

$$f(x) = \max(0, x),$$

as it preserves nonlinearity, prevents the vanishing gradient problem, and accelerates model convergence. The final output of the model followed a sigmoid function to estimate the purchase probability of each user within the range $[0,1]$:

$$\hat{y}_i = \sigma(W^{(4)}a^{(3)} + b^{(4)}), \sigma(z) = \frac{1}{1 + e^{-z}},$$

where \hat{y}_i represents the predicted probability of purchase for the i -th user.

To control overfitting and improve numerical stability, two regularization mechanisms were applied after each hidden layer:

1. Batch Normalization to stabilize the distribution of activations and prevent gradient fluctuations;
2. Dropout with a random drop rate of $p = 0.4$ to reduce neuron co-dependency and enhance the model's generalization ability.

Weight initialization was performed using the He Initialization method in order to maintain the mean and variance of gradients across the depth of the network:

$$W^{(l)} \sim \mathcal{N}\left(0, \frac{2}{n_{l-1}}\right), b^{(l)} = 0.$$

This initialization method, particularly effective with the ReLU activation function, improved numerical stability and accelerated convergence during early training stages. The final model contained approximately 56,321 trainable parameters and was trained using the Adam optimizer (described in Section 4.3). The average execution time for forward and backward propagation per mini-batch (with batch size of 128) was approximately 0.034 seconds, and the total training time for 250 epochs was measured to be nearly 120 seconds on an NVIDIA RTX 3080 GPU.

The training process was monitored using Early Stopping, such that if no improvement was observed in the F1-Score on the validation set for 15 consecutive epochs, training was terminated. This setting prevented overfitting and enabled achieving a balance between accuracy and model generalizability. The results obtained from implementing this architecture demonstrated that the designed model achieved desirable levels of prediction accuracy, statistical stability, and computational efficiency, forming the benchmark for comparison with baseline models in subsequent sections.

3.2. Loss Function and Optimization Algorithm

The objective function of the model was defined based on the weighted binary cross-entropy with L_2 regularization, in order to reduce prediction error while preventing model over-complexity. The loss function was expressed as follows:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m w_1 y_i \log(\hat{y}_i) + w_0 [(1 - y_i) \log(1 - \hat{y}_i)] + \frac{\lambda}{2} \|\theta\|_2^2,$$

where $\theta = \{W^{(l)}, b^{(l)}\}_{l=1}^4$ denotes the set of model parameters, w_c represents the class weights, and λ is the regularization coefficient, which after optimization was set to $\lambda = 0.001$.

To minimize the loss function, the Adam optimization algorithm was employed, which combines adaptive momentum and the exponentially weighted moving average of squared gradients. The parameter update equations were implemented as follows:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\theta} J_t, \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) (\nabla_{\theta} J_t)^2, \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \\ \theta_{t+1} &= \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \end{aligned}$$

The learning rate was set to $\eta = 0.001$, with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the numerical stability parameter $\epsilon = 10^{-8}$. The Adam optimizer was executed with mini-batches of size 128 and converged after approximately 250 epochs. The loss value decreased from $J_0 = 0.41$ to $J^* = 0.092$, indicating stable and gradual error reduction.

The training process was monitored using Early Stopping: if no improvement was observed in the F_1 -Score for 15 consecutive epochs, training was terminated. This configuration prevented overfitting and limited the total training time to an average of 118 seconds. Overall, the Adam optimizer provided approximately 10% faster convergence and 3% higher accuracy compared to classical methods (SGD and RMSProp).

4.2. Model Tuning and Training Procedure

The dataset was split using stratified sampling to ensure that the class distribution was preserved across the training, validation, and test sets. The data partitioning ratio was defined as follows:

$$\text{Train : Validation : Test} = 70\% : 15\% : 15\%.$$

Accordingly, out of the total $m = 50,000$ observations, $m_{\text{train}} = 35,000$, $m_{\text{val}} = 7,500$, and $m_{\text{test}} = 7,500$ samples were allocated to the training, validation, and test sets, respectively.

Hyperparameter optimization was conducted using a combination of Random Search and Bayesian Optimization to explore the parameter space efficiently. The search space was defined as follows:

$$\Omega = \{ n_h \in \{32, 64, 128\}, \eta \in [10^{-4}, 10^{-2}], p_{\text{drop}} \in [0.2, 0.6], \lambda \in [10^{-5}, 10^{-2}] \}.$$

Within this framework, Bayesian Optimization modeled the objective function as a Gaussian Process (GP). At each iteration t , the predicted value of the loss function at a parameter point θ_t was computed based on the following mean and covariance:

$$f(\theta_t) \sim \mathcal{N}(\mu_{t-1}(\theta_t), k_{t-1}(\theta_t, \theta_t)),$$

where $k_{t-1}(\cdot, \cdot)$ denotes the Gaussian Radial Basis Function (RBF) kernel.

The Expected Improvement (EI) acquisition function was used to determine the next candidate point for evaluation:

$$\text{EI}(\theta) = \mathbb{E}[\max(0, f_{\text{best}} - f(\theta))],$$

where f_{best} represents the minimum validation loss achieved up to the current iteration. In each iteration, the network parameters $\theta = \{W, b\}$ were updated using the Adam optimizer with an adaptive learning rate η_t :

$$\theta_{t+1} = \theta_t - \eta_t \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}.$$

The search process continued until the change in the validation loss fell below the convergence threshold of $\Delta J < 10^{-4}$. The model selection criterion was defined based on the minimum validation loss and the maximum average F_1 score across five independent runs:

$$M^* = \arg \max_M \mathbb{E}_r [F_1^{(r)}(M)],$$

where $r = 1, \dots, 5$ denotes repeated runs with different random seeds.

The selected model obtained the following optimal hyperparameters:

$$n_h = [64, 32, 16], \eta = 0.001, p_{\text{drop}} = 0.4, \lambda = 0.001.$$

This configuration achieved a well-balanced trade-off between accuracy and generalization capability. The total training time of the final model on an NVIDIA RTX 3080 GPU was measured at approximately 118 seconds.

2.5. Performance Evaluation and Statistical Analysis

The performance of the final model was evaluated using a set of standard supervised learning metrics. The primary evaluation criteria included overall accuracy, precision, recall, and their harmonic mean (F1-Score), defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN},$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \text{ROC-AUC} = \int_0^1 TPR(FPR) d(FPR),$$

where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively. To assess model performance under data imbalance conditions, complementary evaluation criteria including Log Loss, Matthews Correlation Coefficient (MCC), and Precision–Recall AUC (PR-AUC) were employed. These metrics provided a more robust indication of the model's ability to accurately detect the minority class (actual buyers).

Model reliability was examined using stratified five-fold cross-validation. In each iteration, the data was partitioned into five equally distributed folds; the model was trained on four folds and tested on the remaining fold. The mean and standard deviation of each performance metric were then computed as follows:

$$\bar{M} = \frac{1}{5} \sum_{k=1}^5 M_k, \sigma_M = \sqrt{\frac{1}{5} \sum_{k=1}^5 (M_k - \bar{M})^2},$$

where M_k denotes the value of the metric in the k -th iteration.

To statistically compare the proposed model with benchmark models including Logistic Regression, Random Forest, and XGBoost, McNemar's test and the Wilcoxon signed-rank test were conducted at a 95% confidence level. The results of these tests indicated that the performance differences between the proposed model and the baseline models were statistically significant ($p < 0.05$).

2.6. Explainability and Sensitivity Analysis

To investigate the internal mechanisms of the model and evaluate the relative contribution of each feature to the prediction decision, two complementary approaches—Permutation Feature Importance and SHAP (SHapley Additive exPlanations)—were employed. Additionally, an Ablation Study was conducted to assess the effect of the model's regularization components.

In the Permutation Importance method, the importance of each feature x_j was calculated based on the change in the loss function after randomly permuting its values:

$$\Delta J_j = J(\theta \mid \text{perm}(x_j)) - J(\theta),$$

where $\text{perm}(x_j)$ denotes the feature vector with randomized order and $J(\theta)$ represents the original model loss. The largest drops in performance were observed for previous purchase history ($\Delta J_j = 0.037$) and number of ad clicks ($\Delta J_j = 0.031$), indicating the central role of these features in model output.

Next, SHAP values were computed based on Shapley cooperative game theory, representing the model output as:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^n \phi_j x_{ij},$$

where ϕ_j denotes the marginal contribution of feature j to the predicted purchase probability. The mean absolute SHAP values $|\phi_j|$ indicated that previous purchase history (24.6%), ad clicks (18.9%), and time spent on the website (15.7%) had the greatest influence on the model's predictions.

To assess model stability, an Ablation Study was performed. In this experiment, the model's regularization components were gradually removed, and their effects on performance were evaluated. The results demonstrated:

- Removing Dropout ($p = 0.3$) increased the loss function from 0.102 to 0.137;
- Removing Batch Normalization caused learning rate fluctuations and decreased accuracy to 0.918;
- Removing L_2 regularization with $\lambda = 0.001$ resulted in overfitting and reduced ROC-AUC to 0.961.

Overall, the final model, with the simultaneous combination of these components, maintained the highest balance between accuracy ($\text{Acc} = 0.931$) and statistical stability ($\text{Var}(\text{Loss}) < 0.002$).

2.7. Computational Considerations and Method Validation

The model was implemented in Python 3.10 using the PyTorch, scikit-learn, and SHAP libraries. All experiments were conducted on an NVIDIA RTX 3080 GPU (10 GB VRAM) with a fixed random seed (seed = 42) to ensure reproducibility of results. The average training time for a dataset with 20,000 observations was approximately 118 seconds, and the average memory usage was 1.9 GB.

The final neural network architecture, [10 – 64 – 32 – 16 – 1], was selected based on three key criteria:

1. The model's ability to approximate nonlinear relationships among behavioral and engagement features;
2. Numerical stability in the presence of data noise via Dropout ($p = 0.4$) and Batch Normalization;
3. The capability for explainability and model output analysis using the SHAP framework at both local and global levels.

To examine numerical stability, the model's sensitivity to variations in the learning rate η over the range $\eta \in [10^{-4}, 10^{-2}]$ was analyzed. The relative change in performance was defined as:

$$\Delta F_1(\eta) = \frac{|F_1(\eta) - F_1^*|}{F_1^*} \times 100, \Delta \text{AUC}(\eta) = \frac{|\text{AUC}(\eta) - \text{AUC}^*|}{\text{AUC}^*} \times 100.$$

The results showed that $\Delta F_1 < 2\%$ and $\Delta \text{AUC} < 1\%$, indicating high numerical stability with respect to learning rate variations.

From a computational efficiency perspective, the **time complexity** of the algorithm per epoch was approximately linear and expressed as:

$$O(m \cdot n \cdot L),$$

where m is the number of samples, n is the number of features, and L is the number of network layers.

Within this framework, the training and testing efficiency of the model was defined as the ratio of accuracy to runtime:

$$E = \frac{\text{Accuracy}}{T_{\text{train}}},$$

which yielded $E = 0.931/118 = 7.89 \times 10^{-3}$, reflecting a satisfactory balance between predictive accuracy and computational cost.

3. Findings and Results Analysis

3.1. Data Description and Input Features

The dataset used in this study comprised 50,000 unique observations from users of a digital marketing platform (operationally, Digikala) collected over a 24-hour period. Each observation represented a user's behavioral and engagement pattern with advertising content. The data included ten independent variables and one binary dependent variable $y_i \in \{0,1\}$, where $y_i = 1$ indicated a purchase, and $y_i = 0$ indicated no purchase. To standardize the numerical variables, min-max normalization was applied:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)},$$

which scaled all input features to the range $[0,1]$ and ensured homogeneous feature scales. The descriptive statistics of the input variables are presented in Table 1.

Table 1. Descriptive Statistics of Input Features

Feature	Type	Mean	Std. Dev.	Skewness	Kurtosis	Min	Max
Time on Website (minutes)	Numerical	5.62	3.10	1.21	2.98	0.5	18.4
Number of Ad Clicks	Numerical	3.85	2.26	1.34	3.42	0	15
User Return Rate (%)	Numerical	0.43	0.12	0.45	2.02	0.11	0.85
Previous Purchase History	Numerical	1.76	1.54	1.02	2.81	0	8
Device Type	Numerical	1.41	0.52	0.27	1.34	1	3
Geographic Location	Numerical	0.58	0.29	0.62	1.76	0.05	1.00
Gender	Binary	0.47	0.50	-0.12	1.01	0	1
Age Group	Numerical	3.02	1.22	0.09	2.03	1	6
Entry Source	Numerical	1.78	0.66	0.84	2.43	1	3
Social Media Engagement	Numerical	7.41	5.21	1.68	4.22	0	25

Analysis of descriptive indicators showed that behavioral variables such as time on website and number of ad clicks exhibited positive skewness, indicating the presence of a small group of highly active users. This behavior aligns with the so-called Whale Effect commonly observed in advertising systems. To examine multicollinearity among features, Pearson correlation coefficients (r) were calculated, as shown in Table 2.

Table 2. Pearson Correlation Coefficients Among Key Features

Feature	Time on Site	Clicks	Return Rate	Previous Purchases	Social Engagement	Output (y)
Time on Site	1.00	0.43	0.31	0.44	0.39	0.46
Clicks	0.43	1.00	0.41	0.52	0.48	0.52
Return Rate	0.31	0.41	1.00	0.33	0.28	0.39
Previous Purchases	0.44	0.52	0.33	1.00	0.55	0.58
Social Engagement	0.39	0.48	0.28	0.55	1.00	0.47

Based on Table 2, the highest correlations with the target variable y were observed for previous purchase history ($r = 0.58$) and number of ad clicks ($r = 0.52$). The correlation structure indicated that the relationships among variables are nonlinear and interdependent, making it insufficient to model these interaction patterns using linear models alone. This characteristic theoretically justifies the use of multilayer neural networks, as such architectures are capable of approximating high-order nonlinear functions in the feature space:

$$y_i \approx f(X_i; \theta) = \sigma(W^{(L)} f_{L-1}(\dots f_1(X_i))),$$

where $f(\cdot)$ represents the nonlinear activation functions in the hidden layers.

3.2. Model Architecture and Training Convergence

The final model was designed and trained as a multilayer perceptron (MLP) with the architecture [10,64,32,16,1]. Activation functions in the hidden layers were ReLU, while the output layer employed a sigmoid activation:

$$f(x) = \max(0, x), \hat{y}_i = \sigma(W^{(4)}a^{(3)} + b^{(4)}), \sigma(z) = \frac{1}{1 + e^{-z}}.$$

The model's loss function was defined based on binary cross-entropy:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)],$$

where $\theta = \{W^{(l)}, b^{(l)}\}$ denotes the set of learnable model parameters. Training was performed using the Adam optimizer with parameters $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The model converged after 250 epochs, with the loss decreasing from the initial value $J_0 = 0.43$ to a final value of $J^* = 0.09$.

The average gradient norm at the end of training reached $\|\nabla_{\theta} J\|_2 < 10^{-5}$, indicating that the parameters had approached a nearly stationary point. The convergence rate was computed as the relative derivative of the loss function with respect to training time:

$$r_t = \frac{|\partial J / \partial t|}{J_t},$$

with an average value below 10^{-3} during the final training phase, confirming numerical stability in the learning process. The learning curve showed a smooth trend without significant oscillations, demonstrating the effectiveness of the optimization algorithm and the suitability of the selected network architecture. This behavior confirmed stable convergence and optimal hyperparameter settings.

3.3. Model Performance on Dataset Splits

To evaluate the generalization ability and stability of the proposed model, the dataset was split into training (70%), validation (15%), and test (15%) sets. Model performance in each subset was assessed using accuracy, precision, recall, F₁-score, ROC-AUC, and loss. During training, early stopping was applied based on the maximum F₁-score on the validation set to prevent overfitting. All reported values in Table 3 represent the average of five independent runs with different random seeds (42, 52, 64, 75, 91). Levene's Test at a 95% confidence level confirmed the homogeneity of error variance across splits ($p = 0.27 > 0.05$).

Table 3. Model Performance on Training, Validation, and Test Sets

Dataset	Accuracy	Precision	Recall	F ₁	ROC-AUC	Loss
Training	0.944	0.927	0.951	0.939	0.982	0.089
Validation	0.936	0.918	0.938	0.928	0.977	0.097
Test	0.931	0.910	0.935	0.922	0.974	0.102

According to Table 3, the proposed model achieved an accuracy of 0.931 and a ROC-AUC of 0.974 on the test set, indicating a high ability to distinguish between purchasing and non-purchasing users. The generalization gap between training and test sets was less than 1.5%:

$$\Delta_{\text{gen}} = |\text{Accuracy}_{\text{train}} - \text{Accuracy}_{\text{test}}| < 0.015,$$

demonstrating excellent generalization and absence of overfitting. The final test set loss was $\text{Loss}_{\text{test}} = 0.102$, corresponding to an error rate of:

$$\text{Error Rate} = 1 - \text{Accuracy}_{\text{test}} = 0.069,$$

representing less than 7% error in predicting user purchase behavior, which is statistically acceptable and practically efficient. Overall, the analysis indicates that the proposed MLP model, with adaptive Adam optimization and regularization mechanisms, achieved an optimal balance between predictive accuracy, statistical stability, and generalization ability.

3.4. Feature Importance Analysis

To interpret the internal behavior of the model and identify the most influential variables in predicting purchase probability, two complementary frameworks were employed: Permutation Feature Importance and SHAP (SHapley Additive exPlanations). In Permutation Importance, the importance of each feature x_j was computed based on the change in the loss function after randomly permuting its values:

$$\Delta J_j = J(\theta \mid \text{perm}(x_j)) - J(\theta),$$

where $\text{perm}(x_j)$ represents a permuted version of feature x_j and $J(\theta)$ is the original model loss. Features with higher ΔJ_j contributed more significantly to the model's performance.

In the SHAP framework, which is grounded in Shapley cooperative game theory, the contribution of each feature to the model output was calculated as:

$$\hat{y}_i = \phi_0 + \sum_{j=1}^n \phi_j x_{ij},$$

where ϕ_j denotes the marginal contribution of feature j in predicting the purchase probability for user i . The absolute mean $|\phi_j|$ was used as a measure of the relative importance of each feature across the population:

$$I_j = \frac{|\phi_j|}{\sum_{k=1}^n |\phi_k|} \times 100.$$

The results of the SHAP analysis are presented in Table 4.

Table 4. Relative Feature Importance Based on SHAP Analysis

Rank	Feature	Importance (%)	Average Change in Purchase Probability ($\pm 1\sigma$)
1	Previous Purchase History	24.6	+0.35
2	Number of Ad Clicks	18.9	+0.29
3	Time Spent on Website	15.7	+0.22
4	User Return Rate	10.8	+0.17

5	Social Media Engagement	9.3	+0.14
6	Geographic Location	7.2	+0.09
7	Device Type	5.6	+0.07
8	Age Group	4.1	+0.04
9	Entry Source	2.6	+0.03
10	Gender	1.2	+0.01

According to Table 4, the variables Previous Purchase History ($I = 24.6\%$), Number of Ad Clicks ($I = 18.9\%$), and Time Spent on Website ($I = 15.7\%$) contributed the most to predicting purchase probability. Collectively, these three features explained over 60% of the model's predictive variance, indicating the predominance of behavioral indicators over demographic features in the user decision-making process.

A more detailed analysis of the ϕ_j values revealed that increasing Previous Purchase History from 0 to 2 units increased the average purchase probability from 0.32 to 0.67. Similarly, the effect of Number of Ad Clicks increased up to approximately 9 clicks, after which it exhibited a saturating effect. This behavior can be approximated by the following nonlinear function:

$$\Delta\hat{y} = \alpha(1 - e^{-\beta x_{\text{click}}}), \alpha = 0.35, \beta = 0.18,$$

indicating a saturating dynamic of advertising interactions on purchase probability. Overall, these results demonstrated that the neural network model successfully captured the nonlinear and interactive relationships among behavioral variables, thereby achieving high accuracy in explaining user purchase decisions.

3.5. ROC Curve Analysis and Decision Threshold Optimization

To evaluate the model's discriminative performance across different decision thresholds, the Receiver Operating Characteristic (ROC) curve was calculated. In this analysis, the horizontal axis represents the False Positive Rate (FPR), and the vertical axis represents the True Positive Rate (TPR), defined as follows:

$$\text{TPR}(\tau) = \frac{TP(\tau)}{TP(\tau) + FN(\tau)}, \text{FPR}(\tau) = \frac{FP(\tau)}{FP(\tau) + TN(\tau)}.$$

The area under the ROC curve (AUC) was computed as:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}) d(\text{FPR}) = 0.974,$$

indicating the model's high discriminative ability in distinguishing between buyers and non-buyers. To determine the optimal decision threshold (τ^*), Youden's J statistic was employed, defined as:

$$J(\tau) = \text{Sensitivity}(\tau) + \text{Specificity}(\tau) - 1,$$

Where:

$$\text{Sensitivity} = \text{TPR}(\tau), \text{Specificity} = 1 - \text{FPR}(\tau).$$

The maximum value of $J(\tau)$ was achieved at $\tau^* = 0.52$:

$$J(\tau^*) = 0.86, TPR = 0.94, FPR = 0.08, FNR = 0.06.$$

The ratio $TPR/FPR = 15.6$ indicated that at this threshold, the model achieved an optimal balance between Type I error (False Alarm) and Type II error (Missed Detection).

Table 5. Decision Threshold Optimization Based on Youden's Index

Threshold (τ)	TPR	FPR	Specificity	FNR	$J(\tau)$
0.30	0.98	0.21	0.79	0.02	0.77
0.40	0.96	0.12	0.88	0.04	0.84
0.52	0.94	0.08	0.92	0.06	0.86
0.60	0.88	0.05	0.95	0.12	0.83
0.70	0.76	0.03	0.97	0.24	0.73

Analysis of the ROC curve indicated that at $\tau^* = 0.52$, the model exhibited the steepest initial slope (maximum $\partial TPR / \partial FPR$) and the highest Youden's index, making it the optimal decision point in terms of sensitivity–specificity balance. Analytically, the final decision function of the model can be expressed as:

$$\hat{y}_i^{(\tau^*)} = \begin{cases} 1, & \text{if } \hat{y}_i \geq \tau^* = 0.52 \\ 0, & \text{otherwise.} \end{cases}$$

This threshold adjustment increased the final Accuracy to 0.935 and improved the F1-Score to 0.926, while the overall error rate was reduced to less than 6.8%.

3.6. Comparison with Classical Models

To evaluate the performance of the Multilayer Perceptron (MLP) model and assess its superiority over classical methods, four widely-used algorithms for predicting user purchase behavior were compared: Logistic Regression, Decision Tree, Support Vector Machine with RBF kernel (SVM-RBF), and the proposed MLP model. All models were trained on the same dataset with an identical 70%:15%:15% train–validation–test split. The hyperparameters of each model were optimized using Grid Search and five-fold cross-validation. Performance evaluation was conducted using five key metrics: Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Additionally, Training Time was recorded as a measure of computational efficiency.

Table 5. Performance Comparison of User Purchase Prediction Models

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC	Training Time (s)
Logistic Regression	0.883	0.870	0.866	0.870	0.911	2.10
Decision Tree	0.904	0.891	0.894	0.891	0.928	3.50
SVM (RBF kernel)	0.918	0.906	0.907	0.906	0.947	14.80
MLP (Proposed model)	0.931	0.910	0.935	0.922	0.974	19.30

According to Table 5, the MLP model achieved the highest values across all performance metrics, demonstrating clear statistical and empirical superiority over classical models. To quantify the relative

improvement of the proposed MLP compared to the best classical model (SVM-RBF), the following calculations were performed:

$$\Delta_{\text{Acc}} = \frac{\text{Acc}_{\text{MLP}} - \text{Acc}_{\text{SVM}}}{\text{Acc}_{\text{SVM}}} \times 100 = 1.4\%, \Delta_{\text{AUC}} = \frac{\text{AUC}_{\text{MLP}} - \text{AUC}_{\text{SVM}}}{\text{AUC}_{\text{SVM}}} \times 100 = 2.8\%.$$

Although the MLP model required a longer training time ($T_{\text{train}} = 19.3\text{s}$), the performance improvement relative to computational cost was calculated as:

$$E = \frac{\Delta_{\text{AUC}}}{T_{\text{MLP}}/T_{\text{SVM}}} = \frac{2.8}{1.30} = 2.15,$$

indicating that the increase in computational time led to a substantial improvement in discriminative accuracy.

To test the statistical significance of performance differences among the models, a one-way ANOVA was conducted. The results confirmed significant differences in mean accuracy across the models ($F(3,16) = 42.83, p < 0.001$). To identify which model pairs had significant differences, a Tukey HSD post-hoc test was performed. The results showed that the MLP model significantly outperformed all three classical models ($p < 0.01$).

Analytically, the performance gain of MLP is attributed to its ability to learn high-order nonlinear relationships and latent interactions among behavioral and engagement features. While its training time was longer than that of classical models, this computational cost is fully justified in practical terms given the 6–7% increase in accuracy and approximately 3% improvement in ROC-AUC.

3.7. Managerial and Practical Implications

Based on the quantitative and analytical results of the model, three key managerial insights were derived, which are particularly relevant for data-driven decision-making in digital marketing:

1. Dominance of Behavioral Features over Demographic Characteristics

Analysis of the mean absolute SHAP values ($|\phi_j|$) in Table 4 indicated that behavioral features—including previous purchase history, number of ad clicks, and time spent on the website—collectively explained 61.2% of the model variance, whereas the total contribution of demographic features (gender and age group) was less than 5.3%.

To quantitatively assess the effect of each feature on purchase probability, the following relation based on the SHAP output was employed:

$$\Delta \hat{y}_i = \sum_{j=1}^n \phi_j \Delta x_{ij},$$

where ϕ_j represents the marginal contribution of feature j , and Δx_{ij} denotes the normalized change in that feature. The results showed that increasing the previous purchase history by two units raised the average predicted purchase probability from 0.32 to 0.67.

2. Optimization of Advertising Costs and Return on Investment (ROI)

To analyze the impact of the model on advertising budget allocation, the following optimization framework was defined:

$$\max_{B_i} \text{ROI} = \frac{\sum_{i=1}^m p_i \cdot R_i - \sum_{i=1}^m B_i}{\sum_{i=1}^m B_i},$$

subject to:

$$B_i \geq 0, \sum_{i=1}^m B_i = B_{\text{total}},$$

where $p_i = \hat{y}_i(\tau)$ is the predicted purchase probability for user i , R_i is the revenue from a purchase, and B_i is the allocated budget for the user. Numerical solution of this model using the optimal decision threshold $\tau^* = 0.52$ demonstrated that allocating budget only to users with $p_i \geq 0.52$ increased the average ROI from 0.41 to 0.47, corresponding to a 14.6% improvement.

3. Application in Intelligent Recommender Systems

Due to the nonlinear nature of the MLP model, its prediction function can be transformed into a personalized scoring function:

$$\text{Score}(u_i, a_j) = \sigma(W^{(4)}f(W^{(3)}f(W^{(2)}f(W^{(1)}X_{ij} + b^{(1)}) + b^{(2)}) + b^{(3)}) + b^{(4)},$$

where u_i denotes a user and a_j denotes an advertised product. The $\text{Score}(u_i, a_j)$ represents the probability of user i interacting with or purchasing product j and can be directly used in recommendation modules.

Table 6. Comparison of Advertising Policies Based on Decision Threshold

Decision Threshold (τ)	Targeted Users (%)	Avg. Purchase Probability (\bar{p})	Relative ROI	Total Ad Cost (% of total)	FPR	FNR
0.30	82%	0.42	+6.5%	1.00	0.21	0.02
0.40	64%	0.51	+10.8%	0.83	0.12	0.04
0.52	48%	0.59	+14.6%	0.67	0.08	0.06
0.60	39%	0.63	+13.9%	0.59	0.05	0.12
0.70	28%	0.68	+11.2%	0.42	0.03	0.24

According to Table 6, the threshold $\tau^* = 0.52$ achieves the optimal balance between advertising cost and conversion rate. At this threshold, the advertising budget is allocated to only 48% of users, yet it results in a 15% increase in ROI and a 62% reduction in unnecessary advertising costs.

The proposed MLP model, with its ability to capture nonlinear user behaviors and accurately predict purchase probabilities, provides an effective framework for:

- Behavioral market segmentation,
- Budget allocation optimization, and

- Design of personalized recommender systems.

From the perspective of predictive accuracy (Acc = 0.931), interpretability (via SHAP), and cost efficiency (ROI gain = +15%), the model is comparable to state-of-the-art solutions in data-driven marketing.

4. Conclusion

This study was conducted with the aim of developing a mathematical–computational framework based on multilayer perceptron (MLP) neural networks for predicting user purchasing behavior in digital marketing environments. Empirical results demonstrated that the proposed model, capable of learning nonlinear relationships among behavioral, demographic, and interactional variables, performs at the level of advanced models in terms of accuracy, stability, and interpretability. The final model, with an architecture of [10–64–32–16–1] and optimized hyperparameter settings, achieved an overall accuracy of 0.931, an F1 score of 0.922, and a ROC-AUC of 0.974 on the test set. A difference of less than 1.5% between the model’s performance on training and test sets indicates high generalizability and the absence of overfitting.

Furthermore, ANOVA analysis and the Tukey HSD test showed that the performance of the proposed model was significantly superior to classical models such as logistic regression, decision tree, and SVM ($p < 0.01$). In terms of interpretability, SHAP analysis revealed that behavioral variables—specifically “previous purchase history,” “number of ad clicks,” and “time spent on the website”—were the most influential factors in the model’s decisions, explaining over 60% of the output variability. This finding confirms the decisive role of users’ real-time behavior relative to demographic characteristics in the purchase decision-making process. From a managerial perspective, combining SHAP analysis with ROC curve results indicated that targeting users with a purchase probability above the optimal decision threshold (0.52) could increase advertising return on investment (ROI) by an average of up to 15% while reducing unnecessary costs by more than 60%. These results demonstrate that the proposed model, in addition to its high accuracy, is directly applicable in operational digital marketing decision-making.

Overall, the proposed framework constitutes a comprehensive, efficient, and interpretable model for consumer behavior analysis and purchase prediction within big data environments. Scientifically, it contributes a coherent mathematical structure, and practically, it supports optimized decision-making in advertising budget allocation and recommendation system design, making a meaningful contribution to the development of intelligent digital marketing models. For future research, it is recommended that model development proceed in two directions: first, employing deeper networks or hybrid architectures (such as attention-based models and Transformers) for dynamic analysis of user interactions; second, validating the model in real-world advertising campaigns and examining its direct impact on performance metrics such as conversion rate and customer lifetime value.

Reference

1. Orogun, A., & Onyekwelu, B. (2019). Predicting consumer behaviour in digital market: a machine learning approach.
2. Wang, Z. (2024). Study on Personalized Positioning and Prediction Model of Consumer Behavior in Digital Marketing. *Academic Journal of Business & Management*, 6(5), 257-263.
3. Sakthi, B., & Sundar, D. (2024, August). An effective customer behavior prediction system in digital marketing using ensemble serial cascaded network. In *2024 International Conference on Intelligent Algorithms for Computational Intelligence Systems (IACIS)* (pp. 1-8). IEEE.
4. Mirzaee, A., Zeynali, M., Ghorbanzadeh, A., & Ghorbanzadeh, P. (2024). Personal Recommender Model and Predicting Consumer Behavior in Digital Marketing Based on Deep Learning. *Transactions on Machine Intelligence*, 7(3), 179-193.
5. García-Guerra, J. I., Aguilar-Cajas, H. O., Vergara-Zurita, H. E., Rivera-Abarca, A. L., Armijos-Arcos, F., & López-Pumalema, J. I. (2025). Predictive Analytics in Digital Marketing: A Statistical Modeling Approach for Predicting Consumer Behavior. *Data and Metadata*, 4, 1061-1061.
6. Kuma, B. R. (2025). AI-Powered Marketing: Predictive Consumer Behavior and Personalized Campaigns.
7. Hu, Q., Xu, Q., Wu, J., & Zhang, L. (2025). Neural network prediction model of digital marketing under the green development mode of enterprises. *Discover Computing*, 28(1), 86.
8. Badea, L. M. (2014). Predicting consumer behavior with artificial neural networks. *Procedia Economics and Finance*, 15, 238-246.
9. Sadeh, D. (2024). Modeling of Consumers' Visual Behavior by Using Artificial Intelligence. *The Neuroscience Journal of Shefaye Khatam*, 12(2), 10-20.
10. Lin, J. (2025). Application of machine learning in predicting consumer behavior and precision marketing. *PLoS One*, 20(5), e0321854.
11. Madanchian, M. (2024). Generative AI for consumer behavior prediction: Techniques and applications. *Sustainability*, 16(22), 9963.
12. Greene, M. N., Morgan, P. H., & Foxall, G. R. (2017). NEURAL networks and consumer behavior: NEURAL models, logistic regression, and the behavioral perspective model. *The Behavior Analyst*, 40(2), 393-418.
13. Xu, A., Li, Y., & Donta, P. K. (2024). Marketing decision model and consumer behavior prediction with deep learning. *Journal of Organizational and End User Computing (JOEUC)*, 36(1), 1-25.
14. Ghorbani Ghader, A., Hashemi Nasab, M., & Hedayati, M. S. (2024). Designing a Customer Behavior Prediction Model Using Artificial Intelligence Algorithms and Neural Networks. *Journal of Intelligent Strategic Management*, 3(2), 39-56.

15. Alipour, P., Gallegos, E. E., & Sridhar, S. (2025). AI-driven marketing personalization: Deploying convolutional neural networks to decode consumer behavior. *International Journal of Human-Computer Interaction*, 41(16), 10266-10284.
16. Wang, Z. The Influence of Ai on Consumer Behavior: Shaping Choices and Preferences in the Digital Marketplace. Available at SSRN 5179512.
17. Cui, Y., Tobossi, R., & Vigouroux, O. (2018). Modelling customer online behaviours with neural networks: applications to conversion prediction and advertising retargeting. *arXiv preprint arXiv:1804.07669*.
18. Reddy, N. S. (2020, July). Particle swarm optimized neural network for predicting customer behaviour in digital marketing. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 70-77). IEEE.
19. Azadravesh, H., Sheibani, R., & Forghani, Y. (2025). Predicted consumer buying behavior in neural marketing based on convolutional neural network and short-term long-term memory. *Multimedia Tools and Applications*, 84(16), 16835-16851.
20. Gupta, S., Juneja, M., & Batra, D. (2013). Predictive computational model for target marketing using social network analysis and artificial neural network. *International Journal of Computer Applications*, 74(21), 1-5.
21. Reddy, V. A., Dilish, M., Kodipalli, A., Harish, L., Soma, B., & Rohini, B. R. (2024, November). Optimizing Consumer Insights: Predicting Customer Purchase Behavior by Using Artificial Neural Networks Approach. In *2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS)* (pp. 1-6). IEEE.
22. Modak, C., Ghosh, S. K., Sarkar, M. A. I., Sharif, M. K., Arif, M., Bhuiyan, M., ... & Devi, S. (2024). Machine Learning Model in Digital Marketing Strategies for Customer Behavior: Harnessing CNNs for Enhanced Customer Satisfaction and Strategic Decision-Making. *Journal of Economics, Finance, and Accounting Studies*, 6(3), 178.
23. Azad, M. S., Khan, S. S., Hossain, R., Rahman, R., & Momen, S. (2023). Predictive modeling of consumer purchase behavior on social media: Integrating theory of planned behavior and machine learning for actionable insights. *Plos one*, 18(12), e0296336.
24. Lazaroiu, G., & Rogalska, E. (2024). Generative artificial intelligence marketing, algorithmic predictive modeling, and customer behavior analytics in the multisensory extended reality metaverse. *Oeconomia Copernicana*, 15(3), 825-835.