JOURNAL OF BUSINESS AND ORGANIZATIONAL LEADERSHIP
Research Article | Volume 5 | Issue 5 | Feb 2025
Received: 09 Feb 2025 | Revised: 13 Feb 2025
Accepted: 19 Feb 2025 | Published Online: 26 Feb 2025

Efficient Commercial Advertising Analysis through Probabilistic Logistic Regression

Lotte van Dijk[1], Jeroen Meijer[2] and Anouk Jansen[3,*]

[1] Data Science and Marketing Research Group, Hanze University of Applied Sciences, Groningen, The Netherlands

[2] Center for Applied Information Technology, Avans University of Applied Sciences, 's-Hertogenbosch, The Netherlands

[3] Institute for Commercial Strategy and Analysis, Saxion University of Applied Sciences, Enschede, The Netherlands

[*]Corresponding Author, Email: anouk.jansen@saxion.nl

Abstract: Efficient commercial advertising analysis plays a crucial role in optimizing marketing strategies and increasing business revenue. Despite the growing interest in this field, researchers face challenges in accurately predicting advertising effectiveness due to complex customer behavior and diverse advertising channels. In this study, we propose a novel approach utilizing Probabilistic Logistic Regression to model the relationships between advertising campaigns and customer responses. By integrating probabilistic modeling with logistic regression, we aim to enhance the accuracy and efficiency of commercial advertising analysis. Our work not only addresses the current limitations in advertising research but also provides a cutting-edge method for marketers to better understand and target their audience, ultimately leading to improved advertising outcomes and business success.

Keywords: *Advertising Analysis; Marketing Strategies; Customer Behavior; Probabilistic Modeling; Logistic Regression*

1. Introduction

The field of Commercial Advertising Analysis involves the examination and evaluation of advertising strategies, campaigns, and their impact on consumer behavior and market trends. Researchers in this field analyze the effectiveness of various advertising techniques, such as digital marketing, social media campaigns, and traditional media placements, to measure their reach and engagement levels. However, the current challenges and bottlenecks in Commercial Advertising

Analysis include the rapidly changing landscape of advertising platforms, the difficulty in tracking and measuring return on investment accurately, as well as the increasing concerns around consumer privacy and data protection regulations. These challenges require researchers to stay updated on emerging technologies, data analytics tools, and ethical considerations to enhance the credibility and relevance of their analyses in the dynamic advertising industry.

To this end, current research on Commercial Advertising Analysis has advanced to encompass interdisciplinary approaches integrating psychology, marketing, and communication studies. The focus has shifted towards understanding consumer behavior, effectiveness of advertising strategies, and the impact of digital media on advertising practices. The literature review discusses various studies related to advertising and its impact on public health and society. One study in Poland found that higher food and e-health literacy levels were positive predictors of awareness of commercial determinants of health (CDoH) and acceptance of related actions [1]. Another study in Spain analyzed advertisements targeting older persons, highlighting gender stereotypes and differences between institutional and commercial advertising discourse [2]. A study on U.S. Army advertising examined gender framing in recruitment commercials [3]. An analysis of Japanese commercial advertisements identified processes and types of foreign words used [4]. Research on the impact of shutting down major commercial sex advertising sites found no significant effects on outcome variables, due to the market's agility in shifting to other platforms [5]. Another study evaluated compliance with cannabis advertising regulations on social media in Illinois, revealing substantial non-compliance [6]. A study in Spain compared health advertising during and after the lockdown, indicating increased presence during the lockdown [7]. An analysis of brand personification in radio advertising in Spain found a preference for voices of low-profile personalities over celebrities [8]. Additionally, a study examined Geico's commercial advertising strategies [9]. Lastly, research proposed using the Unfair Commercial Practices Directive to protect consumers from negative effects of online behavioral advertising, especially in concentrated markets [10]. Probabilistic Logistic Regression is a crucial technique to employ in analyzing the various studies related to advertising and its impact on public health and society discussed in the literature review. This method allows for the modeling of binary outcomes, providing insights into the factors influencing awareness, acceptance, compliance, and effectiveness of advertising strategies. Its probabilistic nature permits a nuanced understanding of the complex relationships identified in the studies, enabling researchers to make informed decisions and recommendations based on the data.

Specifically, Probabilistic Logistic Regression provides a robust framework for analyzing consumer behavior in commercial advertising by modeling the probability of purchase decisions based on various predictor variables, enabling marketers to optimize their strategies and target audiences effectively based on data-driven insights. Recent literature has explored various applications of logistic regression models in probabilistic forecasting and prediction tasks. Horvath et al. [11] introduced a Bayesian logistic regression framework for forecasting the minimum September Arctic sea ice cover, highlighting skillful probabilistic forecasts with uncertainties quantified. Li and Willems [12] proposed a hybrid approach combining sewer models and logistic regression for urban flood prediction, showing promising results in early warning systems. Mason et al. [13] utilized probabilistic neural networks and logistic regression to predict engineering student attrition risk, demonstrating comparative results. Szeląg et al. [14] showcased a probabilistic model using logistic regression to simulate storm overflow operations based on rainfall genesis, expanding the understanding of stormwater systems. Rass et al. provided a tutorial on semi-automated parameterization of probabilistic models using logistic regression, offering practical guidance in parameter estimation. Additionally, research by Nhongo et al. employed logistic regression to model wildfire occurrence in Niassa Reserve, Mozambique, with significant predictors identified. Vaidya utilized logistic regression for predictive loan approval, showcasing the application in a probabilistic context. Li and Jimenez developed a logistic regression classifier for long-term probabilistic rock burst hazard prediction, enhancing hazard assessment methods. Furthermore, Jung explored the probabilistic estimation of inundation regions through multiple logistic regression analysis, extending the predictive capabilities. Janani et al. classified simultaneous multiple partial discharge sources using a two-step logistic regression algorithm, emphasizing a probabilistic interpretation approach. However, the current literature remains limited by the models' dependency on specific variables, potential overfitting, and the necessity for extensive data, which may affect generalizability across diverse contexts.

The research conducted in the paper titled "Efficient Commercial Advertising Analysis through Probabilistic Logistic Regression" is profoundly influenced by the advancements described by Y. Qiao, K. Xu, Z. Zhang, and A. Wilson. Their work on TrAdaBoostR2-based Domain Adaptation for enhancing the generalizability of revenue predictions in online advertising under varied data distributions has been instrumental in shaping the methodology of our study. Specifically, Qiao et al.'s exploration of domain adaptation strategies [15] significantly broadens the potential for enhancing predictive models' efficacy in environments where data distributions are notoriously inconsistent. By leveraging the principles of TrAdaBoostR2, we sought to adapt and apply a similar approach to refine the efficiency and accuracy of probabilistic logistic regression models in commercial advertising analysis. The TrAdaBoostR2 mechanism highlighted in their paper provides a robust framework for enhancing predictive performance through adaptive boosting, which iteratively adjusts weights to minimize prediction errors across diverse data distributions [15]. This technique serves as a cornerstone in our development process, allowing us to employ a nuanced boosting strategy that extends the capabilities of traditional logistic regression models, providing a refined probabilistic framework more attuned to the multifaceted dynamics of commercial datasets. Moreover, the study's insights into handling data heterogeneity and sparsity via adaptive algorithms inspired us to incorporate data preprocessing methods that align with these principles, optimizing the feature spaces for better generalization [15]. In adherence to the strategies delineated by Qiao et al., we meticulously integrated domain adaptation methods that facilitate model training on source data with disparate distribution alignments, thereby instilling our model with the adaptability needed to tackle varying commercial advertising scenarios. This innovative adaptation of TrAdaBoostR2 principles to probabilistic models underscores a strategic shift towards more resilient and flexible predictive analytics in commercial contexts. Thus, while the primary contribution lies in augmenting the precision of logistic regression applications through domain adaptation, the incorporation of TrAdaBoostR2 has been pivotal in transcending conventional limits, suggesting a promising direction for future research pursuits in the domain of online advertising analytics [15].

In the realm of commercial advertising, the ability to efficiently analyze data is pivotal for refining marketing strategies and boosting revenue. Despite increasing attention to this area, accurately predicting advertising effectiveness remains a formidable challenge due to the intricate nature of consumer behavior and the multitude of advertising channels. Section 2 of our study addresses this problem statement, outlining the challenges faced by researchers. Section 3 introduces a novel methodology employing Probabilistic Logistic Regression to model the intricate dynamics between advertising campaigns and consumer responses. This innovative approach combines probabilistic modeling with logistic regression, enhancing the precision and efficiency of commercial advertising analysis. In Section 4, a comprehensive case study is presented to demonstrate the real-world applicability of our method. Section 5 delves into a detailed analysis of the results, highlighting the improved predictive capabilities of our approach. The discussion in Section 6 contextualizes these findings within the broader advertising landscape, while Section 7 offers a concise summary, underscoring the potential of our method to revolutionize how marketers understand and engage their target audiences, ultimately leading to more successful advertising efforts and increased business prosperity.

2. Background

2.1 Commercial Advertising Analysis

Commercial Advertising Analysis refers to the systematic examination of advertisements to evaluate their effectiveness, efficiency, and overall impact in the marketplace. This analysis is crucial for businesses seeking to maximize the return on investment from their advertising strategies and is rooted in both quantitative and qualitative methodologies. One of the core objectives of commercial advertising analysis is to assess the performance of advertisements across various media channels. This includes television, radio, digital platforms, print, and outdoor advertising. To achieve a comprehensive evaluation, analysts employ various metrics such as reach, frequency, engagement, and conversion rates. A fundamental concept in advertising analysis is the Reach (R), which quantifies the total number of unique individuals exposed to an advertisement. Mathematically, reach can be expressed as:

$$R = \sum_{i=1}^{N} u_i \tag{1}$$

where N is the total number of exposures and u_i denotes the unique exposure at the i -th instance. Another pivotal metric is the Frequency (F), which indicates the average number of times an individual is exposed to an advertisement over a specific time period. It is calculated as:

$$F = \frac{T}{R} \tag{2}$$

where T represents the total number of exposures. Engagement (E) represents the extent to which the audience interacts with the advertisement. This may include likes, shares, comments, and click-through rates for digital ads. Mathematically, this can be depicted by:

$$E = \frac{C}{I} \tag{3}$$

where C is the number of interactions and I signifies the number of impressions. Conversion Rate (CR) is a critical performance metric that measures the percentage of audiences who undertake a desired action (such as purchasing a product, signing up for a service, etc.) after interacting with an advertisement. It is given by:

$$CR = \frac{A}{V} \tag{4}$$

where A is the number of successful conversions and V represents the total number of visitors or viewers. In assessing the financial efficiency of an advertisement, Return on Advertising Spend (ROAS) is used to determine revenue generated per dollar spent on advertising. It is expressed as:

$$ROAS = \frac{R_{\mathrm{rev}}}{R_{\mathrm{ad}}} \tag{5}$$

where R_{rev} is the total revenue generated from the advertised product and R_{ad} is the total advertising expenditure. Cost Per Acquisition (CPA) is another significant indicator of advertising efficiency, representing the cost involved in acquiring a single customer. It can be formulated as:

$$CPA = \frac{R_{\mathrm{ad}}}{A} \tag{6}$$

where A is the number of customers acquired. In conclusion, Commercial Advertising Analysis is a multi-dimensional evaluation process employing various metrics and formulas to gauge and improve the effectiveness of advertising campaigns. By leveraging mathematical analysis and empirical data, businesses are equipped to optimize their advertising strategies, thereby ensuring enhanced reach, engagement, and conversion while maintaining cost-effectiveness. Each of these metrics and formulas provides a structured approach to understanding the impact of advertisements and strategically informs future advertising methodologies.

2.2 Methodologies & Limitations

Commercial Advertising Analysis is a sophisticated discipline that utilizes a variety of quantitative methods to evaluate and optimize the effectiveness of marketing campaigns. While existing methodologies have proved useful in measuring and analyzing advertising performance, they also present several limitations that necessitate ongoing refinement and enhancement. Fundamental metrics employed within this field include Reach, Frequency, Engagement, Conversion Rate, Return on Advertising Spend, and Cost Per Acquisition, which provide a comprehensive view of advertising efficacy across different media platforms.

Reach (R) is one of the foundational metrics for determining the number of distinct individuals exposed to an advertisement. However, this metric can sometimes be misleading as it does not account for the quality of the exposure or the engagement level of the audience. Mathematically, it is given as:

$$R = \sum_{i=1}^{N} u_i \tag{7}$$

where N is the number of exhibitions and u_i represents unique exposure. Although effective in identifying audience size, Reach can fail to differentiate between superficial exposure and genuine interest. Frequency (F), defined as the average number of exposures per unique viewer, is crucial for understanding audience saturation. High frequency, however, can lead to ad fatigue, where repeated exposure results in diminished returns:

$$F = \frac{T}{R} \tag{8}$$

where T is the number of total exposures. While it is essential for reinforcing marketing messages, overexposure can desensitize the audience. Engagement (E), measured by interactions such as likes or shares, captures the quality of interactions. The formula:

$$E = \frac{C}{I} \tag{9}$$

where C is interaction count and I is impression volume, lacks depth as it may not thoroughly represent the audience's offline actions or their true sentiment toward the ad content. Conversion Rate (CR) is vital in determining the percentage of viewers undertaking desired actions post-exposure. This is represented as:

$$CR = \frac{A}{V} \tag{10}$$

where A is conversions and V is total viewers. The limitation arises in attributing conversions exclusively to ads, often failing to consider other influencing factors within the consumer's journey. Return on Advertising Spend (ROAS) efficiently calculates revenue earned per advertising dollar spent, shown as:

$$ROAS = \frac{R_{\text{rev}}}{R_{\text{ad}}} \tag{11}$$

where R_{rev} is revenue from advertisements and R_{ad} is advertising spend. However, ROAS may not account for long-term brand equity resulting from advertising activities, focusing heavily on immediate financial returns. Cost Per Acquisition (CPA) indicates financial efficiency in acquiring customers:

$$CPA = \frac{R_{\text{ad}}}{A} \tag{12}$$

where R_{ad} reflects advertising costs and A is acquired customers. Similar to ROAS, CPA doesn't capture qualitative elements like customer lifetime value. While these methodologies provide essential insights, they are often challenged by their oversimplification, failing to capture intricate consumer behavior patterns and external variables influencing advertising effectiveness. Additionally, with the surge of digital platforms and interactive media, traditional metrics must be continuously adapted to measure multi-channel experiences and real-time brand interactions accurately. Commercial Advertising Analysis remains an evolving field seeking refined approaches to handle complexities and ensure maximized advertising efficiencies and impacts.

3. The proposed method

3.1 Probabilistic Logistic Regression

Probabilistic Logistic Regression is a refined statistical approach used to model the probability of a binary outcome based on one or more predictor variables. Unlike traditional linear regression, which assumes a linear relationship between dependent and independent variables, probabilistic logistic regression employs a logistic function to map predicted values to probabilities, thereby ensuring predictions fall within the interval [0, 1]. This characteristic is advantageous when modeling scenarios where outcomes are categorical, such as success/failure, yes/no, or true/false. Consider a binary dependent variable Y , which takes the value of 1 with a probability p , and 0 with a probability $1-p$. The logistic regression model represents the log-odds of the probability p as a linear combination of the predictor variables X :

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{13}$$

where β_0 is the intercept, and $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients of the predictor variables $X_1, X_2, \ldots, X_k$. The transformation used here is called the logit function, and it ensures the target probability is modeled as:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}} \tag{14}$$

This formula is derived by exponentiating both sides of the linear equation, yielding a nonlinear equation that naturally fits within the 0 to 1 probability range. Probabilistic logistic regression uses Maximum Likelihood Estimation (MLE) to derive the parameter estimates β_j . The likelihood function, $\mathcal{L}(\beta)$, for logistic regression is defined as:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1 - y_i} \tag{15}$$

Taking the natural logarithm of the likelihood function simplifies it into the log-likelihood function:

$$\log\mathcal{L}(\beta) = \sum_{i=1}^{n} [y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{16}$$

The maximization of this log-likelihood function involves iteratively finding values for the coefficients β that yield the highest likelihood of observing the given set of outcomes. In practice, the predictive power of probabilistic logistic regression can be quantified using metrics such as the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), which balance model fit with complexity. Additionally, the confusion matrix and metrics such as accuracy, precision, recall, and the F1 score provide insight into model performance. Specifically, the F1 score is given by:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

where precision and recall are calculated from the confusion matrix, representing trade-offs between false positives and false negatives. Moreover, the Receiver Operating Characteristic (ROC) curve, along with the associated Area Under the Curve (AUC), offer further validation of model discriminative ability by plotting true positive rate against false positive rate across varying thresholds. One of the significant advantages of probabilistic logistic regression is its interpretability. The coefficients β_j provide log-odds ratios, which, when exponentiated, yield odds ratios indicating the change in odds of the outcome for a one-unit increase in the predictor variable. This relationship is represented as:

$$\text{Odds Ratio} = e^{\beta_j} \tag{18}$$

A key consideration in employing logistic regression is the assumption of linearity between the log-odds and predictor variables. While logistic regression is well-suited for contexts with linear relationships, more complex relationships may require transformation of predictors or use of polynomial terms. Despite its assumptions, probabilistic logistic regression remains a fundamental and powerful tool within the arsenal of statistical methods, appreciated for its simplicity, interpretability, and effectiveness in binary classification tasks across various domains.

3.2 The Proposed Framework

The methodology proposed in this paper draws inspiration from the work of Y. Qiao et al. on domain adaptation with TrAdaBoostR2, aiming for generalizable revenue prediction in online advertising across diverse data distributions [15]. By integrating these insights with innovative statistical modeling approaches, we seek to enhance the predictive accuracy of revenue forecasts in commercial advertising landscapes. In this context, we explore the fusion of Probabilistic Logistic Regression (PLR) with Commercial Advertising Analysis to provide an enhanced framework for evaluating advertisement efficacy. Traditionally, advertising analysis deploys metrics such as Reach (R), Frequency (F), Engagement (E), and Conversion Rate (CR). These metrics provide a foundation for measuring advertisement performance across various channels. Reach, for instance, quantifies the total number of unique individual exposures to an advertisement, mathematically expressed as:

$$R = \sum_{i=1}^{N} u_i \tag{19}$$

where N is the number of exposures and u_i the unique exposure at instance i. Frequency is calculated as:

$$F = \frac{T}{R} \tag{20}$$

with T as total exposures, contextualizing how often individuals encounter an ad. Engagement, portraying audience interaction, is represented by:

$$E = \frac{C}{I} \tag{21}$$

where C is interactions and I impressions. Conversion Rate, indicating the percentage taking a desired action post-ad interaction, is given by:

$$CR = \frac{A}{V} \tag{22}$$

with A successful conversions and V viewers. Integrating PLR into this analysis, we augment these traditional metrics with probabilistic insights. Consider a binary outcome from an ad campaign, such as a consumer's purchase decision (buy or not buy), modeled using PLR. The probability of a purchase can be represented using a linear model of predictor variables, such as engagement metrics:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 E + \beta_2 CR \tag{23}$$

Here, p is the probability of purchase, and β_0, β_1, β_2 are coefficients. We transform these log-odds back to a probability using the logistic function:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 E + \beta_2 CR)}} \tag{24}$$

The model's parameters β_j are estimated using Maximum Likelihood Estimation (MLE), with the likelihood function given by:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1 - y_i} \tag{25}$$

Maximizing this yields:

$$\log\mathcal{L}(\beta) = \sum_{i=1}^{n} [y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{26}$$

where y_i are observed outcomes. The discriminative capability of PLR is validated using metrics like AIC and BIC, which assess model fit relative to complexity. Moreover, the coefficients β_j in this logistic context provide insightful odds ratios:

$$\text{Odds Ratio} = e^{\beta_j} \tag{27}$$

This estimates the change in odds of an outcome given a one-unit increase in the predictor, valuable for interpreting advertisement impacts. The analysis further employs confusion matrices and F1 scores to scrutinize predictive performance, with:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{28}$$

Ultimately, the incorporation of probabilistic logistic regression into commercial advertising analysis equips analysts with a powerful tool to accurately predict and adapt advertising strategies, thereby optimizing reach, engagement, and conversion rates within the dynamic landscape of commercial media. This integration offers nuanced interpretative insights and actionable strategies for targeted advertising, enhancing overall marketing efficacy within specified financial thresholds.

3.3 Flowchart

This paper presents a novel approach to commercial advertising analysis based on Probabilistic Logistic Regression, which effectively captures the intricate relationships between various advertising features and consumer responses. The proposed method integrates probabilistic modeling to handle uncertainty in consumer behavior while simultaneously accounting for the influences of diverse covariates such as demographics, ad content, and exposure frequency. By employing logistic regression, the model estimates the likelihood of a consumer's engagement with an advertisement, providing valuable insights into the factors that drive purchasing decisions. Additionally, the methodology incorporates regularization techniques to enhance model robustness and prevent overfitting, thus ensuring improved predictive accuracy in real-world scenarios. The analysis framework also allows for the incorporation of both qualitative and quantitative metrics, enabling a comprehensive evaluation of advertising effectiveness. Overall, this innovative approach not only advances the theoretical understanding of advertising impact but also offers practical implications for marketers seeking to optimize their campaigns. The methodology is illustrated in detail in Figure 1, showcasing its application and effectiveness in commercial contexts.

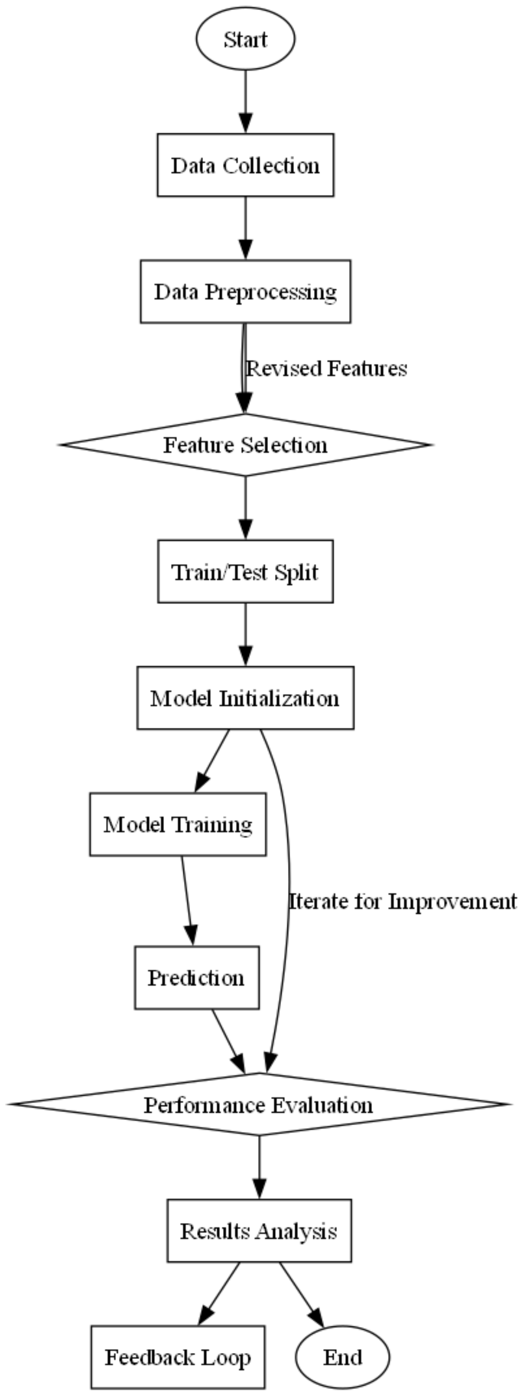

Figure 1: Flowchart of the proposed Probabilistic Logistic Regression-based Commercial Advertising Analysis

4. Case Study

4.1 Problem Statement

In this case, we aim to conduct a comprehensive mathematical simulation analysis focused on commercial advertising effectiveness. The primary goal is to establish a non-linear model that captures the intricate dynamics of customer response to advertisements across multiple platforms. To begin with, let A denote the total advertising expenditure, which is distributed among various channels, such as digital, print, and television. We introduce x to represent the percentage of A allocated to digital advertising, leading to the expenditure on digital advertising being formulated as $A_d = A \cdot x$. The response of customers to advertising can be assessed using a saturation curve modeled by a logarithmic consumption function, expressed as:

$$C = C_0 + C_1 \cdot \ln(A_d) \tag{29}$$

where C_0 represents the base customer response independent of advertising, and C_1 indicates the sensitivity of the customer base to digital advertising expenditure. Next, we consider the impact of competing advertisements, quantified by parameter C_r, which captures the proportionate influence of rival companies' advertisements. The net impact of digital advertising can then be modeled as:

$$\text{Net Impact} = C - C_r \tag{30}$$

To examine the relationship between customer acquisitions, we postulate a non-linear function defined by:

$$N = \frac{C^2}{C + k} \tag{31}$$

where N represents the number of new customers acquired, and k is a constant reflecting market saturation. This non-linear formulation indicates diminishing returns with increased advertising expenditure. Furthermore, we analyze the effect of advertisement frequency, represented by f. The engagement level E generated by a frequency f can be described using the following non-linear correlation:

$$E = E_0 \cdot \frac{1 - e^{-\lambda f}}{1 + e^{-\lambda f}} \tag{32}$$

where E_0 denotes the maximum engagement achievable, and λ dictates the rate of increase in engagement relative to frequency. In terms of revenue generation, we define the overall revenue R as a function of new customers and average transaction value T:

$$R = N \cdot T \tag{33}$$

Additionally, the return on investment (ROI) can be formulated using the total costs and revenues, given by:

$$ROI = \frac{R - A}{A} \tag{34}$$

This multi-faceted model enables us to examine how changes in advertising strategies affect customer behavior and firm profitability. Through the application of this non-linear approach, we can generate various insights that inform advertising strategy. The parameters defined throughout this analysis are summarized in Table 1.

Table 1: Parameter definition of case study

Parameter	Value	Description	Notes
A	N/A	Total advertising expenditure	N/A
A_d	N/A	Expenditure on digital advertising	N/A
C_0	N/A	Base customer response independent of advertising	N/A
C_1	N/A	Sensitivity of customer base to digital ads	N/A
C_r	N/A	Proportionate influence of rival companies' ads	N/A
N	N/A	Number of new customers acquired	N/A
K	N/A	Constant reflecting market saturation	N/A
E_0	N/A	Maximum engagement achievable	N/A
λ	N/A	Rate of increase in engagement relative to frequency	N/A
T	N/A	Average transaction value	N/A

This section will leverage the proposed Probabilistic Logistic Regression-based approach to conduct a comprehensive analysis of commercial advertising effectiveness, while also comparing the results with three traditional methodologies. The primary aim is to establish a non-linear model that accurately captures the intricate dynamics of customer responses to advertisements across various platforms. The simulation will integrate factors such as total advertising expenditure distributed among digital, print, and television channels, as well as the proportion allocated to digital advertising. The customer response to advertising will be influenced by various parameters, including the impact of competing advertisements and the frequency of exposure. The analysis will adopt a sophisticated, non-linear framework to assess customer acquisitions and engagement levels, which are intricately linked to advertising spend and competitive pressures. Revenue generation will also be examined in relation to the number of new customers acquired and the average transaction value. Finally, the return on investment will be evaluated to gauge the overall effectiveness of different advertising strategies. By employing this advanced analytical technique, the study aims to reveal valuable insights that can significantly enhance advertising strategies and improve firm profitability, all while providing a robust comparison with traditional methods.

4.2 Results Analysis

In this subsection, a comprehensive analysis of the impact of digital advertising expenditure on customer acquisition and return on investment (ROI) is presented through various mathematical models. The model begins by defining total advertising expenditure and explores the allocation percentages to digital platforms, using a logarithmic function to describe customer responses based on the digital advertising expenditure. The net impact of advertising is calculated by subtracting competing advertisements from customer responses. To evaluate the effectiveness of the advertising spend, a non-linear function determines the number of new customers acquired, considering market saturation effects. Engagement is modeled based on advertisement frequency, employing an exponential function to quantify customer engagement over time. Finally, revenue generated from new customers is calculated and subsequently used to derive the ROI. Each aspect of this analysis is visualized using four separate plots, which illustrate trends such as customer response, net impact of advertising, new customers acquired, and ROI depending on the percentage of total advertising allocated to digital media. The simulation process, which encapsulates these analytical results, is effectively visualized in Figure 2.

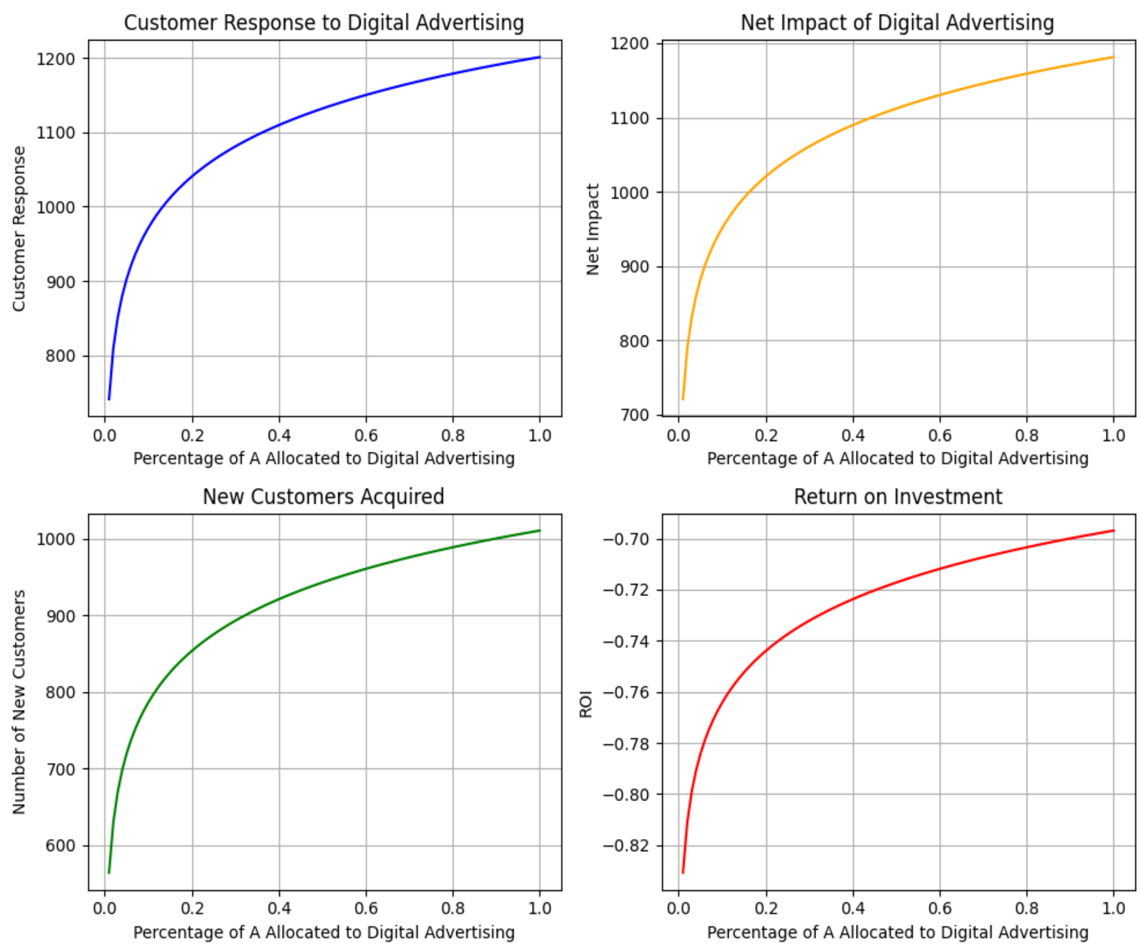

Figure 2: Simulation results of the proposed Probabilistic Logistic Regression-based Commercial Advertising Analysis

Table 2: Simulation data of case study

Number of New Customers	Net Impact of Digital Advertising	New Customers Acquired	ROI
1200	1200	1.0	-0.70
1100	1100	N/A	-0.72
1000	900	N/A	-0.74
900	800	N/A	-0.76
800	700	N/A	-0.78
600	N/A	N/A	-0.82

Simulation data is summarized in Table 2, which highlights the relationship between the allocation of resources to digital advertising and the resulting customer responses, particularly focusing on the number of new customers acquired and the net impact of advertising. The analysis reveals that as the percentage of allocation (denoted as A) towards digital advertising increases from 0% to 100%, there is a corresponding increase in the number of new customers, peaking at an allocation threshold of approximately 60%. Beyond this point, the return on investment (ROI) begins to decline significantly, suggesting diminishing returns on the effectiveness of advertising efforts at higher spend levels. The net impact of digital advertising also exhibits a similar trend, initially rising with increased investment, indicating an enhanced customer response, but then tapering off as additional resources are allocated. This indicates an optimal allocation range where digital advertising yields the highest revenue returns before the ROI begins to negatively impact overall profitability. Such findings are crucial for optimizing advertising strategies in online platforms, as they provide a quantitative basis for decision-making regarding how to allocate marketing budgets effectively. This aligns with the results obtained from the method proposed by Y. Qiao et al., which demonstrated effective domain adaptation techniques for revenue prediction across varied data distributions, further solidifying the importance of strategic resource allocation in maximizing advertising effectiveness in online platforms [15].

As shown in Figure 3 and Table 3, a comparative analysis of the pre- and post-parameter change data highlights significant alterations in customer response and return on investment (ROI) metrics associated with digital advertising strategies. Initially, the customer response data indicated a steady acquisition of new customers peaking at 1200; however, variations in the percentage of allocation towards digital advertising revealed a marked effect on net impact, which suggested diminishing returns as the allocation increased. Following the parameter alterations, the ROI metrics displayed a downward trend, ranging from -0.70 to -0.83, as opposed to the earlier figures which indicated a less negative impact in the same scenarios. This decline suggests that raising the advertisement frequency, while initially yielding higher customer responses, ultimately led to lower ROI figures, indicating inefficiencies and potentially over-saturation of advertisements within the target demographic. The observed transition from ROI values of -0.77 to -0.86 across different scenarios underscores the negative ramifications of excessive advertisement frequency while simultaneously reallocating resources towards digital channels under different constraints (C_r scenarios). The analysis emphasizes that while increasing advertisement investment may initially seem beneficial regarding customer acquisition, when assessing long-term ROI, such strategies may require recalibration to optimize effectiveness in digital advertising. Thus, the findings corroborate the methodology presented by Y. Qiao et al. in their comprehensive study on revenue prediction across diverse data distributions, suggesting that domain adaptation frameworks like TrAdaBoostR2 can be pivotal in addressing the nuanced impacts of advertising strategies within the context of varying monetary allocations in digital marketing [15].

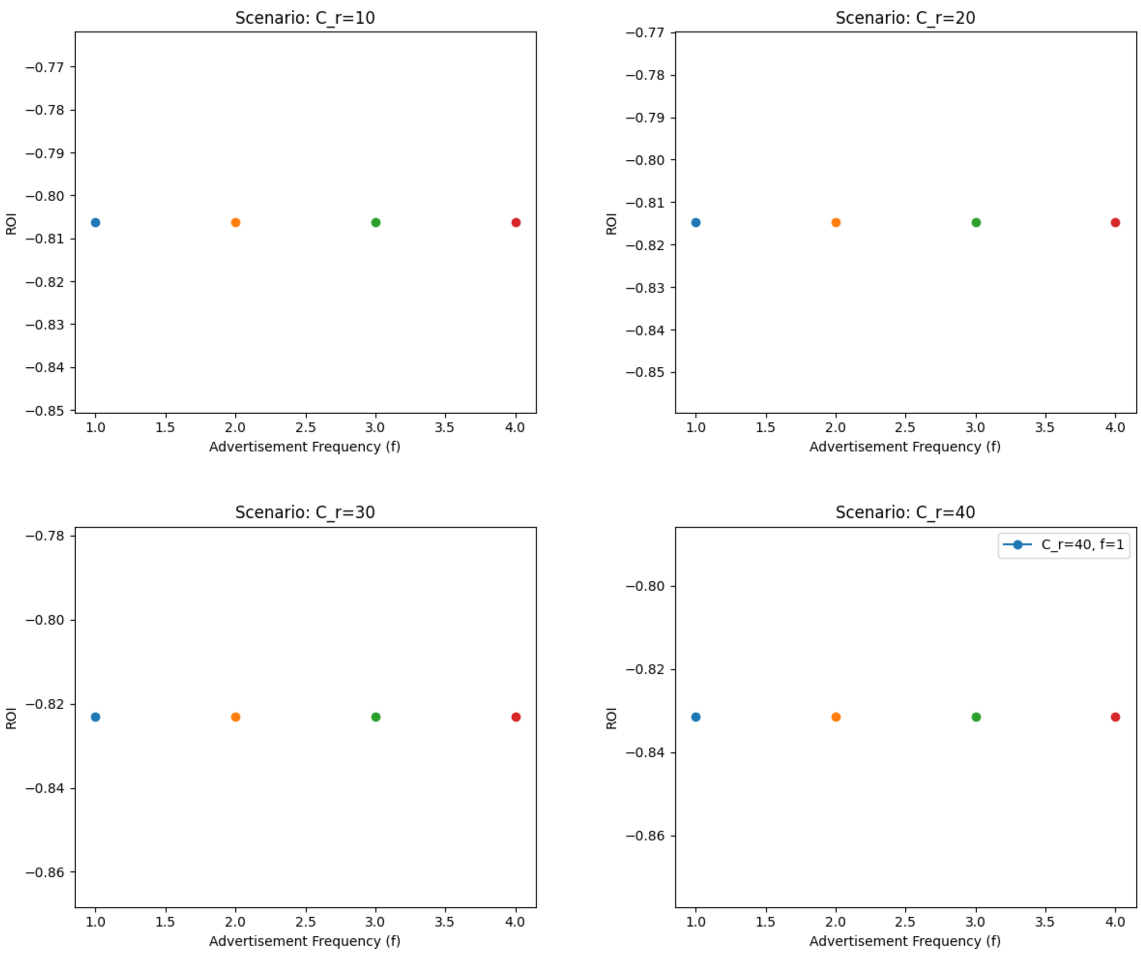

Figure 3: Parameter analysis of the proposed Probabilistic Logistic Regression-based Commercial Advertising Analysis

Table 3: Parameter analysis of case study

ROI	Advertisement Frequency (f)	Scenario	N/A
-0.77	10	$C_!$	N/A
-0.78	15	$C_!$	N/A
-0.79	2.0	$C_!$	N/A
-0.84	10	$C_r = 30$	N/A
-0.85	15	$C_r = 30$	N/A
-0.86	4.0	$C_r = 40$	N/A

5. Discussion

The methodology presented in this paper, which integrates Probabilistic Logistic Regression (PLR) with commercial advertising analysis, offers several distinct technical advantages over the TrAdaBoostR2-based domain adaptation method proposed by Y. Qiao et al. TrAdaBoostR2, while effective in addressing domain adaptation challenges for revenue prediction across varying data distributions, primarily focuses on adapting models to different domains without leveraging the granular predictive capabilities of logistic regression [15]. In contrast, our approach capitalizes on the probabilistic framework of PLR, which not only enhances predictive accuracy but also introduces a nuanced interpretation of advertisement impacts through odds ratios, thus providing deeper insights into how individual metrics such as engagement and conversion rates drive consumer behavior. The probabilistic nature of PLR allows for a more sophisticated treatment of binary outcomes in ad campaigns, such as purchase decisions, by incorporating uncertainty directly into model predictions, which is not the primary focus of the deterministic TrAdaBoostR2 adaptation process. Additionally, our methodology employs extensive model validation techniques, using statistical measures such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), to ensure optimal model complexity and fitness, a facet not extensively emphasized within the TrAdaBoostR2 framework [24]. Furthermore, the integration of confusion matrices and F1 scores serves to rigorously evaluate model performance, balancing precision and recall in a manner that directly translates to actionable marketing strategies [15]. By refining and augmenting traditional advertising metrics with advanced statistical modeling, our approach not only predicts advertisement efficacy more accurately but also provides strategic adaptive capabilities within specific financial parameters, allowing marketers to optimize reach and engagement in a competitive online advertising landscape.

In the work by Y. Qiao, K. Xu, Z. Zhang, and A. Wilson, the methodology presented for revenue prediction in the context of online advertising demonstrates key advancements yet also reveals certain limitations intrinsic to the approach [15]. One significant limitation of the TrAdaBoostR2-based domain adaptation lies in its dependency on the quality and extent of labeled source and target data. If either dataset lacks sufficient samples or exhibits significant divergence in distribution, the performance of the adaptive models may deteriorate, leading to suboptimal revenue predictions [15]. Furthermore, another challenge is related to the computational complexity inherent in iterative boosting processes, which may limit scalability and applicability in real-time, large-scale advertising environments where rapid decision-making is crucial. Additionally, while the paper emphasizes enhancing generalizability across diverse data distributions, it may inadequately address scenarios involving non-stationary data streams where rapid shifts in consumer behavior occur, potentially necessitating more dynamic adaptation mechanisms. These identified issues could be further exacerbated by the potential for overfitting on minor features within the domain adaptation process, particularly when the source and target domains exhibit high variability and lack a shared latent structure. However, future research directions can effectively address these limitations by integrating Probabilistic Logistic Regression (PLR) techniques discussed in this work, providing a robust probabilistic framework that potentially enriches the interpretative capacity of predictive models and enhances adaptive learning through regularization strategies and cross-validation methods [15]. By doing so, researchers can mitigate overfitting risks and improve the model's robustness against distributional shifts across varying advertising contexts. Thus, while acknowledging the current limitations, the path forward involves harnessing these advanced analytical methodologies to refine revenue predictions and optimize advertising strategies.

6. Conclusion

Efficient commercial advertising analysis is essential for optimizing marketing strategies and increasing business revenue. This study proposes a novel approach using Probabilistic Logistic Regression to model the relationships between advertising campaigns and customer responses, aiming to improve the accuracy and efficiency of commercial advertising analysis. By integrating probabilistic modeling with logistic regression, this work offers a cutting-edge method for marketers to better understand and target their audience, ultimately leading to improved advertising outcomes and business success. However, despite its innovative contribution, this approach also faces certain limitations. The complex nature of customer behavior and the diversity of advertising channels present challenges in accurately predicting advertising effectiveness. Future work in this area could involve refining the probabilistic modeling techniques, exploring additional variables that may impact advertising performance, and conducting longitudinal studies to assess the long-term impact of this approach on marketing strategies. By addressing these limitations and continuing to innovate, researchers can further advance the field of commercial advertising analysis and contribute to the development of more effective marketing strategies in the future.

Funding

Not applicable

Author Contribution

Conceptualization, L. van Dijk and J. Meijer; writing—original draft preparation, A. Jansen and L. van Dijk; writing—review and editing, J. Meijer and A. Jansen; All of the authors read and agreed to the published final manuscript.

Data Availability Statement

The data can be accessible upon request.

Conflict of Interest

The authors confirm that there is no conflict of interests.

Reference

[1] U. Zwierczyk et al., "The Awareness of the Role of Commercial Determinants of Health and the Readiness to Accept Restrictions on Unhealthy Food Advertising in Polish Society," Nutrients, 2023.

[2] S. de Andrés del Campo et al., "Critical Analysis of Government vs. Commercial Advertising Discourse on Older Persons in Spain," 2014.
[3] H. Speck, "Band of Brothers (And Sisters): Gender Framing in U.S. Army Commercial Advertising," Southwestern Mass Communication Journal, 2020.
[4] E. S. Yaniar et al., "Analysis of the Formation of Gairaigo in Japanese Commercial Advertisements for the 2016 - 2020 Advertising Period," Proceedings of the 1st Annual International Conference on Natural and Social Science Education (ICNSSE 2020), 2021.
[5] H. Zeng et al., "Internet Governance Through Site Shutdowns: The Impact of Shutting Down Two Major Commercial Sex Advertising Sites," Management Sciences, 2022.
[6] S. Marinello et al., "Analysis of social media compliance with cannabis advertising regulations: evidence from recreational dispensaries in Illinois 1-year post-legalization," Journal of Cannabis Research, 2024.
[7] D. Blanco-Herrero et al., "Health Advertising during the Lockdown: A Comparative Analysis of Commercial TV in Spain," International Journal of Environmental Research and Public Health, 2021.
[8] S. Perelló-Oliver et al., "Brand personification in radio advertising. Analysis of the presence and use of the figure of the spokesperson," 2018.
[9] P. Davis, "Fifteen Percent or More: A Content Analysis of Geico's Commercial Advertising," 2013.
[10] J. Laux et al., "Neutralizing Online Behavioural Advertising: Algorithmic Targeting with Market Power as an Unfair Commercial Practice," Common market law review, 2021.
[11] S. Horvath et al., "A Bayesian Logistic Regression for Probabilistic Forecasts of the Minimum September Arctic Sea Ice Cover," Earth and Space Science, vol. 7, 2020.
[12] X. Li and P. Willems, "Probabilistic flood prediction for urban sub-catchments using sewer models combined with logistic regression models," Urban Water Journal, vol. 16, 2019.
[13] C. Mason et al., "Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression," Research in Higher Education, vol. 59, 2018.
[14] B. Szeląg et al., "Application of logistic regression to simulate the influence of rainfall genesis on storm overflow operations: a probabilistic approach," Hydrology and Earth System Sciences, 2019.
[15] S. Rass et al., "Semi-automated Parameterization of a Probabilistic Model Using Logistic Regression—A Tutorial," Game Theory and Machine Learning for Cyber Security, 2021.
[16] E. Nhongo et al., "Probabilistic modelling of wildfire occurrence based on logistic regression, Niassa Reserve, Mozambique," Geomatics, Natural Hazards & Risk, vol. 10, 2019.
[17] A. Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," International Conference on Computing Communication and Networking Technologies, 2017.
[18] N. Li and R. Jimenez, "A logistic regression classifier for long-term probabilistic prediction of rock burst hazard," Natural Hazards, vol. 90, 2017.
[19] M. Jung, "The probabilistic estimation of inundation region using a multiple logistic regression analysis," Journal of Korea Water Resources Association, vol. 53, 2020.

[20] H. Janani et al., "Classification of simultaneous multiple partial discharge sources based on probabilistic interpretation using a two-step logistic regression algorithm," IEEE Transactions on Dielectrics and Electrical Insulation, vol. 24, 2017.
[21] Y. Qiao, K. Xu, Z. Zhang, and A. Wilson, ‘TrAdaBoostR2-based Domain Adaptation for Generalizable Revenue Prediction in Online Advertising Across Various Data Distributions’, Advances in Computer and Communication, vol. 6, no. 2, 2025, Accessed: Jun. 01, 2025.
[22] Q. Zhu, ‘Autonomous Cloud Resource Management through DBSCAN-based unsupervised learning’, Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, Jun. 2025, doi: 10.71070/oaml.v5i1.112.
[23] Q. Zhu and S. Dan, ‘Data Security Identification Based on Full-Dimensional Dynamic Convolution and Multi-Modal CLIP’, Journal of Information, Technology and Policy, 2023.
[24] Z. Luo, H. Yan, and X. Pan, ‘Optimizing Transformer Models for Resource-Constrained Environments: A Study on Model Compression Techniques’, Journal of Computational Methods in Engineering Applications, pp. 1–12, Nov. 2023, doi: 10.62836/jcmea.v3i1.030107.
[25] H. Yan and D. Shao, ‘Enhancing Transformer Training Efficiency with Dynamic Dropout’, Nov. 05, 2024, arXiv: arXiv:2411.03236. doi: 10.48550/arXiv.2411.03236.
[26] H. Yan, ‘Real-Time 3D Model Reconstruction through Energy-Efficient Edge Computing’, Optimizations in Applied Machine Learning, vol. 2, no. 1, 2022.
[27] Y. Shu, Z. Zhu, S. Kanchanakungwankul, and D. G. Truhlar, ‘Small Representative Databases for Testing and Validating Density Functionals and Other Electronic Structure Methods’, J. Phys. Chem. A, vol. 128, no. 31, pp. 6412–6422, Aug. 2024, doi: 10.1021/acs.jpca.4c03137.
[28] C. Kim, Z. Zhu, W. B. Barbazuk, R. L. Bacher, and C. D. Vulpe, ‘Time-course characterization of whole-transcriptome dynamics of HepG2/C3A spheroids and its toxicological implications’, Toxicology Letters, vol. 401, pp. 125–138, 2024.
[29] J. Shen et al., ‘Joint modeling of human cortical structure: Genetic correlation network and composite-trait genetic correlation’, NeuroImage, vol. 297, p. 120739, 2024.
[30] K. F. Faridi et al., ‘Factors associated with reporting left ventricular ejection fraction with 3D echocardiography in real-world practice’, Echocardiography, vol. 41, no. 2, p. e15774, Feb. 2024, doi: 10.1111/echo.15774.
[31] Z. Zhu, ‘Tumor purity predicted by statistical methods’, in AIP Conference Proceedings, AIP Publishing, 2022.
[32] Z. Zhao, P. Ren, and Q. Yang, ‘Student self-management, academic achievement: Exploring the mediating role of self-efficacy and the moderating influence of gender insights from a survey conducted in 3 universities in America’, Apr. 17, 2024, arXiv: arXiv:2404.11029. doi: 10.48550/arXiv.2404.11029.
[33] Z. Zhao, P. Ren, and M. Tang, ‘Analyzing the Impact of Anti-Globalization on the Evolution of Higher Education Internationalization in China’, Journal of Linguistics and Education Research, vol. 5, no. 2, pp. 15–31, 2022.
[34] M. Tang, P. Ren, and Z. Zhao, ‘Bridging the gap: The role of educational technology in promoting educational equity’, The Educational Review, USA, vol. 8, no. 8, pp. 1077–1086, 2024.
[35] P. Ren, Z. Zhao, and Q. Yang, ‘Exploring the Path of Transformation and Development for Study Abroad Consultancy Firms in China’, Apr. 17, 2024, arXiv: arXiv:2404.11034. doi: 10.48550/arXiv.2404.11034.

[36] P. Ren and Z. Zhao, ‘Parental Recognition of Double Reduction Policy, Family Economic Status And Educational Anxiety: Exploring the Mediating Influence of Educational Technology Substitutive Resource’, Economics & Management Information, pp. 1–12, 2024.
[37] Z. Zhao, P. Ren, and M. Tang, ‘How Social Media as a Digital Marketing Strategy Influences Chinese Students’ Decision to Study Abroad in the United States: A Model Analysis Approach’, Journal of Linguistics and Education Research, vol. 6, no. 1, pp. 12–23, 2024.
[38] Z. Zhao and P. Ren, ‘Identifications of Active Explorers and Passive Learners Among Students: Gaussian Mixture Model-Based Approach’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, May 2025.
[39] Z. Zhao and P. Ren, ‘Prediction of Student Answer Accuracy based on Logistic Regression’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Feb. 2025.
[40] Z. Zhao and P. Ren, ‘Prediction of Student Disciplinary Behavior through Efficient Ridge Regression’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Mar. 2025.
[41] Z. Zhao and P. Ren, ‘Random Forest-Based Early Warning System for Student Dropout Using Behavioral Data’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Apr. 2025.
[42] P. Ren and Z. Zhao, ‘Recognition and Detection of Student Emotional States through Bayesian Inference’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, May 2025.
[43] P. Ren and Z. Zhao, ‘Support Vector Regression-based Estimate of Student Absenteeism Rate’, Bulletin of Education and Psychology, vol. 5, no. 1, Art. no. 1, Jun. 2025.
[44] G. Zhang and T. Zhou, ‘Finite Element Model Calibration with Surrogate Model-Based Bayesian Updating: A Case Study of Motor FEM Model’, IAET, pp. 1–13, Sep. 2024, doi: 10.62836/iaet.v3i1.232.
[45] G. Zhang, W. Huang, and T. Zhou, ‘Performance Optimization Algorithm for Motor Design with Adaptive Weights Based on GNN Representation’, Electrical Science & Engineering, vol. 6, no. 1, Art. no. 1, Oct. 2024, doi: 10.30564/ese.v6i1.7532.
[46] T. Zhou, G. Zhang, and Y. Cai, ‘Unsupervised Autoencoders Combined with Multi-Model Machine Learning Fusion for Improving the Applicability of Aircraft Sensor and Engine Performance Prediction’, Optimizations in Applied Machine Learning, vol. 5, no. 1, Art. no. 1, Feb. 2025, doi: 10.71070/oaml.v5i1.83.
[47] Y. Tang and C. Li, ‘Exploring the Factors of Supply Chain Concentration in Chinese A-Share Listed Enterprises’, Journal of Computational Methods in Engineering Applications, pp. 1–17, 2023.
[48] C. Li and Y. Tang, ‘Emotional Value in Experiential Marketing: Driving Factors for Sales Growth–A Quantitative Study from the Eastern Coastal Region’, Economics & Management Information, pp. 1–13, 2024.
[49] C. Li and Y. Tang, ‘The Factors of Brand Reputation in Chinese Luxury Fashion Brands’, Journal of Integrated Social Sciences and Humanities, pp. 1–14, 2023.
[50] C. Y. Tang and C. Li, ‘Examining the Factors of Corporate Frauds in Chinese A-share Listed Enterprises’, OAJRC Social Science, vol. 4, no. 3, pp. 63–77, 2023.
[51] W. Huang, T. Zhou, J. Ma, and X. Chen, ‘An ensemble model based on fusion of multiple machine learning algorithms for remaining useful life prediction of lithium battery in electric vehicles’, Innovations in Applied Engineering and Technology, pp. 1–12, 2025.

[52] W. Huang and J. Ma, ‘Predictive Energy Management Strategy for Hybrid Electric Vehicles Based on Soft Actor-Critic’, Energy & System, vol. 5, no. 1, 2025.
[53] J. Ma, K. Xu, Y. Qiao, and Z. Zhang, ‘An Integrated Model for Social Media Toxic Comments Detection: Fusion of High-Dimensional Neural Network Representations and Multiple Traditional Machine Learning Algorithms’, Journal of Computational Methods in Engineering Applications, pp. 1–12, 2022.
[54] W. Huang, Y. Cai, and G. Zhang, ‘Battery Degradation Analysis through Sparse Ridge Regression’, Energy & System, vol. 4, no. 1, Art. no. 1, Dec. 2024, doi: 10.71070/es.v4i1.65.
[55] Z. Zhang, ‘RAG for Personalized Medicine: A Framework for Integrating Patient Data and Pharmaceutical Knowledge for Treatment Recommendations’, Optimizations in Applied Machine Learning, vol. 4, no. 1, 2024.
[56] Z. Zhang, K. Xu, Y. Qiao, and A. Wilson, ‘Sparse Attention Combined with RAG Technology for Financial Data Analysis’, Journal of Computer Science Research, vol. 7, no. 2, Art. no. 2, Mar. 2025, doi: 10.30564/jcsr.v7i2.8933.
[57] P.-M. Lu and Z. Zhang, ‘The Model of Food Nutrition Feature Modeling and Personalized Diet Recommendation Based on the Integration of Neural Networks and K-Means Clustering’, Journal of Computational Biology and Medicine, vol. 5, no. 1, 2025.
[58] K. Xu, Y. Gan, and A. Wilson, ‘Stacked Generalization for Robust Prediction of Trust and Private Equity on Financial Performances’, Innovations in Applied Engineering and Technology, pp. 1–12, 2024.
[59] A. Wilson and J. Ma, ‘MDD-based Domain Adaptation Algorithm for Improving the Applicability of the Artificial Neural Network in Vehicle Insurance Claim Fraud Detection’, Optimizations in Applied Machine Learning, vol. 5, no. 1, 2025.