

Article

Predictive Modeling of Customer Response to Marketing Campaigns

Mohammed El-Hajj ^{*,†}  and Miglena Pavlova [†] 

Department of Semantics, Cybersecurity & Services, University of Twente, 7522 NB Enschede, The Netherlands; za.megy@gmail.com

* Correspondence: m.elhajj@utwente.nl

† These authors contributed equally to this work.

Abstract: In today's data-driven marketing landscape, predicting customer responses to marketing campaigns is essential for optimizing both engagement and Return On Investment (ROI). This study aims to develop a predictive model using a Decision Tree (DT) to identify key factors influencing customer behavior and improve campaign targeting. The methodology involves building the DT model, initially achieving an accuracy of 87.3%. However, the model faced challenges with precision and recall due to class imbalance. To address this, a resampling technique was applied, which significantly improved model performance, increasing recall from 44% to 83.1% and the F1-score from 49% to 74.2%. Key influential features identified include the recency of a customer's purchase, their duration as a customer, and their response history to previous campaigns. This study demonstrates the practicality and interpretability of the DT model, offering actionable insights for marketing professionals seeking to enhance campaign effectiveness and customer targeting.

Keywords: customer relationship management; customer response prediction; decision tree model; F1-score; ROI optimization; predictive modeling



Citation: El-Hajj, M.; Pavlova, M. Predictive Modeling of Customer Response to Marketing Campaigns. *Electronics* **2024**, *13*, 3953. <https://doi.org/10.3390/electronics13193953>

Academic Editor: Domenico Ursino

Received: 18 September 2024

Revised: 2 October 2024

Accepted: 6 October 2024

Published: 7 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In today's highly competitive and data-driven business environment, understanding customer behavior is crucial for effective marketing strategies [1]. Predicting how customers will respond to marketing campaigns not only improves the effectiveness of these efforts but also significantly increases ROI [2]. Despite the vast amount of customer data available, transforming this data into actionable insights remains a major challenge for businesses. The complexity of customer behavior, influenced by factors such as demographics, purchase history, and engagement with previous campaigns, adds to this difficulty.

Traditional marketing strategies often rely on broad, generalized approaches, which lack the precision needed for personalized marketing. This study highlights the increasing need for data-driven decision-making, particularly in predicting customer responses to targeted marketing campaigns. The inability to leverage customer data effectively leads to inefficient resource allocation and missed opportunities for engagement. The primary problem this research addresses is the gap between data collection and the practical implementation of predictive models for improved marketing strategies. Marketing campaigns are a fundamental strategy used by businesses to promote products, services, or brand messages to a target audience. They consist of coordinated marketing efforts that focus on reaching specific goals within a defined timeframe. To further clarify the role of marketing campaigns, we emphasize that they are dynamic efforts aimed at increasing sales, improving customer retention, and fostering brand loyalty. The outcome of these campaigns can be measured through metrics such as response rates, sales conversion, and customer feedback. Understanding the intricacies of marketing campaigns allows us to better address the challenge of predicting customer responses, which is the central focus of this work.

Predictive modeling, particularly using Decision Tree (DT) models, offers a promising solution by using historical data to forecast future customer behavior [3]. DT models provide interpretability and ease of use, making them valuable tools for marketers looking to understand which factors most influence customer responses. However, one of the significant limitations encountered in this research is the class imbalance in the customer response data, which can skew model results and reduce the effectiveness of predictions.

Marketing strategies can generally be divided into mass marketing and direct marketing. Mass marketing uses widespread media platforms like television and radio to reach existing and potential customers. In contrast, direct marketing focuses on contacting specific clients directly, often proving more cost-effective and resource-efficient. Understanding the effectiveness of these strategies requires a deep understanding of customer behavior. Raorane and Kulkarni [4] suggest that studying consumer psychology, mindset, behavior, and motivation allows companies to refine their marketing strategies. Therefore, collecting and analyzing customer data is essential for businesses.

Customer Relationship Management (CRM) systems facilitate the automatic collection of customer data, such as demographics, purchase history, and interactions with the company. These data allow businesses to make informed decisions and tailor their marketing efforts. Traditionally, customer behavior prediction relied heavily on intuition and experience, often based on general trends rather than precise, data-driven insights. However, with the rise of Machine Learning (ML), more sophisticated and accurate models have emerged.

Tree-based ML classifiers, such as DT and Random Forest (RF) models, are known for their high accuracy and interpretability. DT models are particularly favored for their simplicity, as they create a tree-like structure of decisions based on input features [3]. RF models, on the other hand, are an ensemble method that improves the predictive power of DT by aggregating the results of multiple trees, improving generalization, and reducing overfitting [5]. Although RF is more robust against noisy data compared to a single DT [5,6], when interpretability is a key requirement and the dataset is relatively small, DT models are preferable. Their simple decision rules make it easier for stakeholders to understand how features contribute to predictions [7].

Despite the potential of predictive modeling, businesses often face significant difficulties in accurately predicting customer responses to marketing campaigns. The complexity of customer behavior, influenced by many factors such as demographics and past interactions, makes it difficult to develop reliable models. Traditional approaches tend to overlook these complexities, leading to generalized and less effective marketing strategies [8].

This study seeks to address this gap by focusing on the interpretability and explainability of the predictive model, utilizing the DT algorithm [3]. The primary objective is to identify and understand the most influential demographic factors, such as age, income, marital status, and education level, as well as to examine the impact of past interactions with the company, including previous purchases and engagement with earlier campaigns. In this article, we explore marketing campaigns in the context of customer engagement and response prediction. Marketing campaigns typically involve a combination of advertising, promotions, public relations, and direct marketing strategies. Key examples include email marketing, social media campaigns, digital advertising, and personalized promotions based on customer data. For instance, a company may launch an email campaign targeting customers who have previously shown interest in specific product categories. These campaigns often use segmentation and targeting methods to optimize effectiveness and customer reach. To address these objectives, the research is guided by the following questions:

- RQ1 What are the challenges and limitations presented in the literature regarding predicting customer marketing responses?
- RQ2 How effective is the DT model at predicting customer response to marketing campaigns?
- RQ3 What are the key factors influencing customer response to marketing campaigns as identified by the DT model?

- Which demographic factors are most influential in predicting customer response to marketing campaigns according to the DT model?
- How do past interactions with the company affect future responses according to the DT model?

The rest of the paper is organized as follows: Section 2 reviews related work, discussing existing literature and the performance of DTs in predictive analytics. The methodology and practical implementation are detailed in Section 3, while Section 4 presents the research findings. Section 5 discusses the results and their implications, and Section 6 concludes with a summary of key findings, limitations, and directions for future research.

2. Related Works

This section reviews key studies that investigate the application of various predictive models in direct marketing, highlighting their methodologies and results. For clarity, we have categorized the studies based on the predictive models used: Decision Trees (DT), Random Forests (RF), and other machine learning techniques such as Support Vector Machine (SVM), Neural Network (NN), Logistic Regression (LR), and Naive Bayes (NB).

2.1. Decision Trees and Their Variants

Decision Trees (DTs) have been extensively applied in marketing campaigns due to their interpretability and practical insights. For instance, Sérgio Moro [9] applied different data mining algorithms such as Naive Bayes, DT, and SVM, and found that SVM had the highest prediction performance. However, DTs were identified as valuable for providing insights into customer attributes that impact marketing outcomes. Another study by Choi et al. [10] demonstrated the effectiveness of DT models in forecasting customer responses, achieving 87.23% accuracy in predicting non-responders.

Furthermore, Liu and Yang [11] explored the application of the DT algorithm and compared it with SVM, NN, NB, and RF. DT outperformed other models with an accuracy of 91.35%, highlighting its superiority in content marketing prediction. Safarkhani and Sérgio Moro [12] utilized the J48 DT algorithm to predict telemarketing success, achieving an accuracy of 94.39%. DT models consistently deliver high performance while maintaining interpretability, making them ideal for marketing applications where transparency is important.

2.2. Random Forest

Random Forest (RF), an ensemble of DTs, has been proposed to enhance predictive accuracy by overcoming the limitations of individual DTs. Asare-Frempong and Jayabalan [13] compared RF, DT, LR, and MLPNN models in predicting customer responses to bank direct marketing campaigns. RF achieved the highest accuracy (86.8%), outperforming DT (84.7%), while DT provided better interpretability. In contrast, Apampa [14] found that RF did not consistently improve the performance of DT for bank marketing response prediction, suggesting that while RF is powerful, it may not always be the best option depending on the context.

2.3. Other Machine Learning Models

SVM, NN, and LR are popular models for direct marketing predictions, often compared to DTs and RFs. K. Wisaeng [15] compared SVM, J48-graft, LAD tree, and RBFN, with SVM achieving the highest accuracy (86.95%). Sérgio Moro et al. [16] applied NN, LR, DT, and SVM to a Portuguese bank dataset, with NN performing best, achieving a success rate of 79% in customer classification.

Olson et al. [17] compared traditional RFM models with advanced techniques such as DT, NN, and LR. DT and NN outperformed the RFM model, showing an optimal balance between accuracy and interpretability. RF also performed well but lacked the same level of interpretability offered by DT. Table 1 highlights various predictive models used in customer response prediction studies, showcasing the performance of Decision Trees, Random Forests, and other models across different contexts, with Decision Trees frequently achieving the highest accuracy and providing interpretability in several cases.

Table 1. Benchmark of related works in predictive models for direct marketing.

Study	Models Compared	Best Accuracy Achieved	Key Findings
K. Wisaeng [15]	SVM, J48-graft, LAD tree, RBFN	86.95% (SVM)	SVM outperformed other models, RBFN had the lowest accuracy.
Sérgio Moro et al. [16]	Logistic regression, neural networks, decision tree, SVM	79% (Neural networks)	Neural networks performed best in predicting customer behavior; targeting top half of customers improved outcomes.
Sérgio Moro [9]	Naive Bayes, decision tree, SVM	N/A (SVM)	SVM showed the highest prediction performance, with call duration as the most significant feature.
Choi et al. [10]	Decision tree	87.23% (Non-responders)	Decision tree accurately predicted non-responders but had lower accuracy for predicting responders.
Usman-Hamza et al. [18]	Decision tree, naive Bayes, K-nearest neighbor	93.6% (Decision tree)	Decision tree outperformed other models in customer churn analysis for live stream e-commerce.
Chaubey et al. [1]	Random forest, decision tree	N/A (Random forest)	Random forest performed better than decision tree for churn prediction but lacks interpretability.
Olson et al. [17]	RFM, logistic regression, decision tree, neural networks	N/A (Decision tree)	Advanced models like decision trees and neural networks outperformed the traditional RFM model. Decision trees offered the best balance between accuracy and interpretability.
Liu & Yang [11]	Decision tree, SVM, neural network, naive Bayes, random forest	91.35% (Decision tree)	Decision tree achieved the highest accuracy in content marketing prediction; precision and F1 score also favored the decision tree model.
Safarkhani & Moro [12]	Decision tree, naive Bayes, logistic regression	94.39% (Decision tree)	Decision tree outperformed other models, achieving the highest accuracy in predicting customer behavior for bank telemarketing.
Asare-Frempong & Jayabalan [13]	Random forest, decision tree, logistic regression, MLPNN	86.8% (Random forest)	Random forest outperformed other models, while decision tree provided useful interpretability for customer term deposit subscription prediction.
Apampa [14]	Random forest, decision tree	N/A (Decision tree)	Random forest did not consistently improve decision tree's performance in bank marketing prediction, favoring decision tree for interpretability.

3. Proposed Solution

This research follows a six-stage methodology that is designed to be straightforward and interpretable for individuals with a moderate understanding of data mining. The whole procedure is shown in Figure 1.

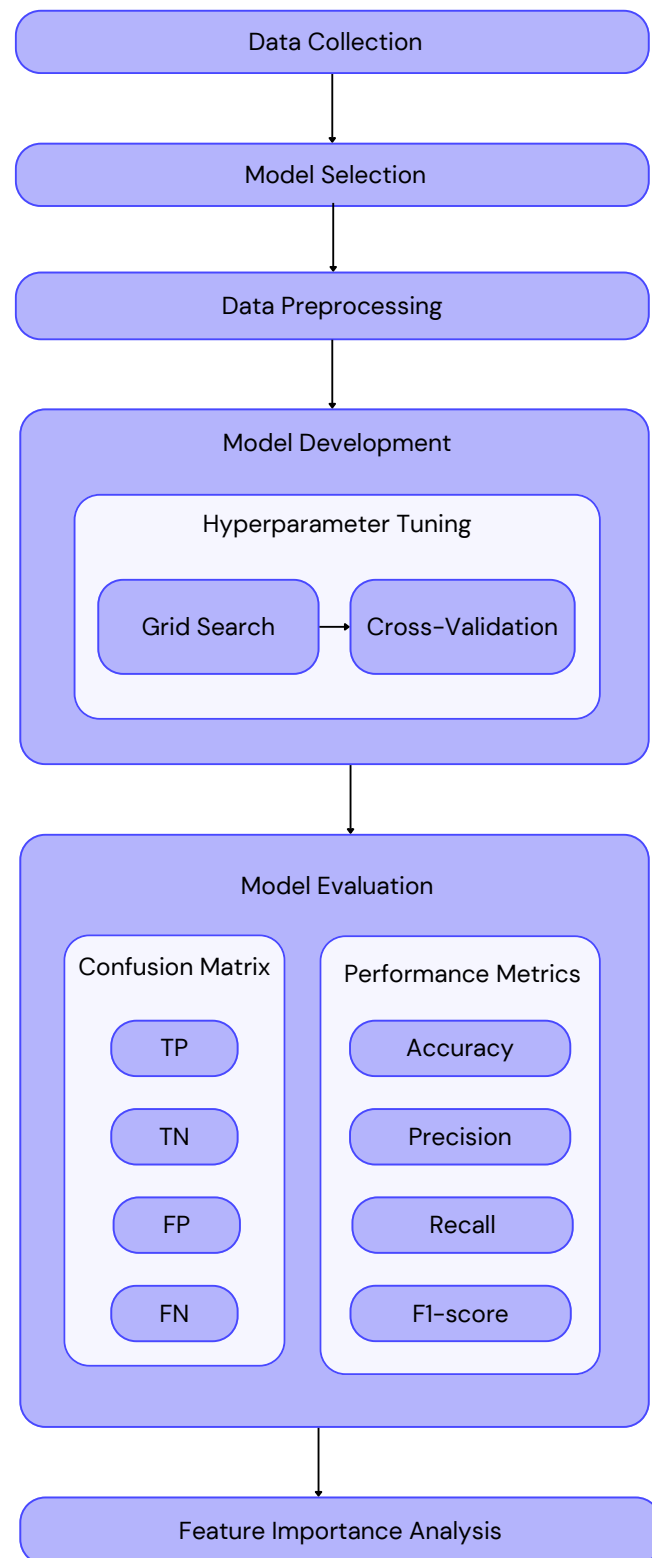


Figure 1. Sketch of the proposed solution.

3.1. Research Design

The research design for this study is structured to effectively address the research questions and validate the proposed solutions. This section outlines the research design employed in this study, emphasizing the methodology utilized to address the challenges identified in the literature and to evaluate the effectiveness of the Decision Tree (DT) model in predicting customer responses to marketing campaigns.

To begin, a comprehensive literature review was conducted to identify the prevailing challenges and limitations in predicting customer marketing responses. This review revealed a critical issue: class imbalance within the dataset. Addressing this challenge, the study employed resampling techniques to mitigate the imbalance, significantly enhancing the model's predictive performance. The results presented in Section 4, demonstrate the model's initial struggles with a higher number of false negatives and a low recall rate, confirming the necessity of the applied resampling method.

Next, the effectiveness of the DT model was evaluated using performance metrics that include accuracy, precision, recall, and F1-score. Initial results indicated high accuracy; however, the other metrics reflected the model's limitations prior to resampling. Following the implementation of undersampling techniques, substantial improvements were observed, particularly in recall and F1-score, affirming the model's enhanced ability to predict positive responses effectively.

Furthermore, a feature importance analysis was conducted to ascertain the key factors influencing customer responses. This analysis identified demographic factors, such as age and income, along with past interactions, as significant predictors of marketing campaign success. By evaluating these features, the research design highlights the importance of customer characteristics and engagement history, providing valuable insights into effective marketing strategies.

In summary, the research design incorporates a systematic approach to addressing class imbalance, evaluating model performance, and identifying influential factors, ultimately leading to improved predictive capabilities and insights into customer behavior.

3.2. Hardware and Software Configuration

The hardware and software configuration for this research ensures the reproducibility of the experiment. In Table 2 the specific components and tools used are listed.

Table 2. Hardware and Software Configurations.

		Component	Configuration
Hardware		Processor	Intel Core i7-10510U
		RAM	16 GB
		Storage	952 GB
		OS	Windows 11 Pro
Software		Language	Python
		Libraries	pandas, seaborn, matplotlib, scikit-learn
		Environment	Jupyter Notebook

To ensure transparency and accessibility in the research process, the source code for this research is made publicly available in a GitHub repository (<https://github.com/megi2002/Predictive-modeling-of-Customer-Response-to-Marketing-Campaigns>, accessed on 4 October 2024).

3.3. Data Collection

The dataset used in this study was obtained from the online platform Kaggle and it belongs to the Brazilian food ordering platform iFood [19]. As presented in Table 3, it

includes various demographic data, such as age, income, marital status, and education level, as well as customer interaction data, such as previous purchases and previous marketing responses. The total number of instances is 2206. The dataset consists of 39 attributes, with the target variable 'Response' as a binary indicator. This target variable has two classes, "yes" indicating that the customer responded positively to a marketing campaign, and "no" indicating that the customer responded negatively. Notably, the dataset contains no categorical data. All attributes are either numerical or binary indicators. This structure eliminates the need for encoding categorical variables.

Table 3. Data dictionary [20].

Demographic	Income	Kidhome	Age
	Teenhome	Customer_Days	marital_Together
	marital_Single	marital_Divorced	marital_Widow
	education_PhD	education_Master	education_Graduation
	education_Basic	education_2n Cycle	
Customer Interaction	MntWines	MntFruits	MntGoldProds
	MntMeatProducts	MntFishProducts	MntSweetProducts
	NumStorePurchases	NumCatalogPurchases	NumWebVisitsMonth
	NumDealsPurchases	NumWebPurchases	Recency
	Z_CostContact	Z_Revenue	MntTotal
	MntRegularProds	Complain	Response
	AcceptedCmp1	AcceptedCmp2	AcceptedCmp3
	AcceptedCmp4	AcceptedCmp5	AcceptedCmpOverall

However, the dataset has a significant class imbalance of the target variable, as visualized in Figure 2. Of the 2206 total instances, 1872 customers did not respond positively to the marketing campaign ('No'), while only 333 responded positively ('Yes'). This results in a ratio of approximately 5.6:1, meaning the majority class vastly outweighs the minority class. The dataset also contains interesting insights into customer demographics, particularly income and purchase behavior. As seen in Figure 3, the income distribution is right-skewed, with a majority of customers earning between USD 50,000 and USD 80,000. This concentration suggests that iFood's customer base primarily consists of middle- to upper-middle-income earners. Notably, few customers have incomes below USD 30,000 or above USD 90,000, suggesting that iFood's marketing and services mainly target middle-income groups. Similarly, the distribution of recency, shown in Figure 4, indicates that a majority of customers made a purchase 20 to 40 days before the campaign. This suggests that iFood's customers typically make relatively frequent purchases, with recency gradually decreasing beyond the 40-day mark. The class imbalance in the dataset highlights the need for proper data preprocessing. Without addressing this imbalance, model performance may be compromised.

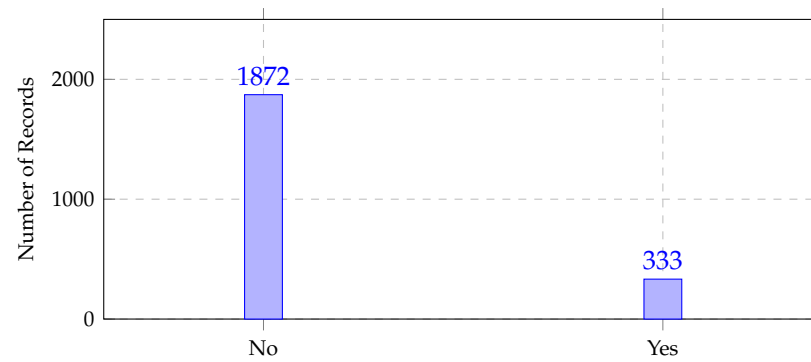


Figure 2. Class distribution of the Response variable showing a significant imbalance [20].

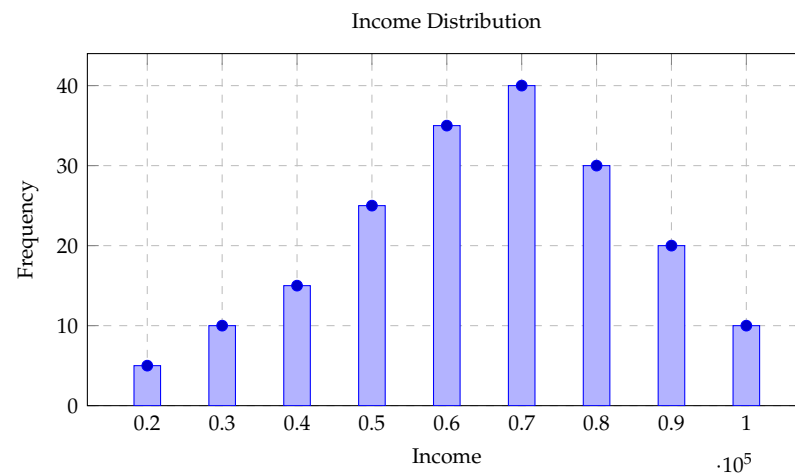


Figure 3. Income distribution [20].

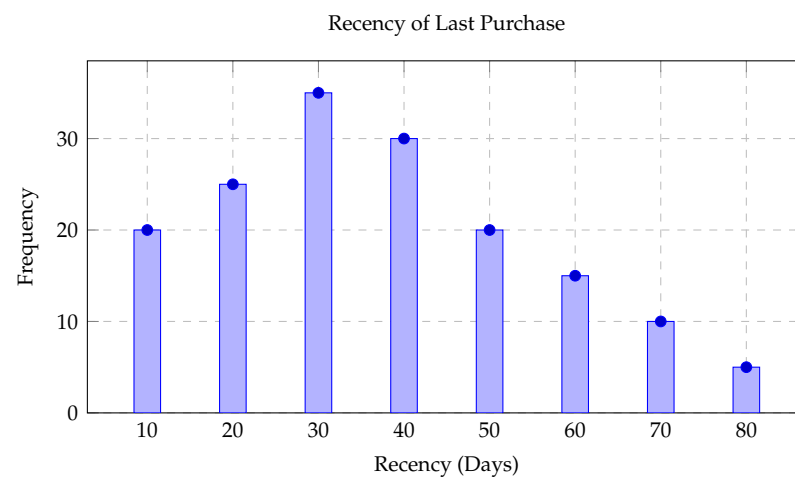


Figure 4. Recency of last purchase [20].

3.4. Model Selection

The Decision Tree (DT) is a supervised machine learning method designed to establish a relationship between input features and the target variable for accurate predictions [3]. Structurally, decision trees resemble a tree where each node signifies a decision based on an attribute, each branch corresponds to an outcome of that decision, and each leaf node represents a target class label. The classification process involves tracing a path from the root node, the primary attribute, to a leaf node [3].

This intuitive method utilizes an “if-else” logic, making it straightforward to understand and interpret [3,7]. This characteristic is particularly beneficial in marketing, where

decisions are often made by individuals with limited technical knowledge. To enhance interpretability further, pruning techniques were employed to reduce the complexity of the tree while maintaining accuracy, which also mitigates the risk of overfitting.

While other techniques, such as Support Vector Machines (SVMs), Neural Networks (NNs), or Random Forest (RF) ensembles, might offer higher accuracy in certain scenarios, they often come with greater complexity and lower transparency. In contrast, the DT model provides a favorable balance between ease of interpretation and reasonable predictive performance. This makes it an ideal choice for marketing applications, where understanding the motivation behind predictions is as important as the accuracy itself.

The trade-off between interpretability and performance is a key reason why DT was selected for this study. Specifically, the model's ability to reveal the significance of various input features aids in identifying key drivers of customer responses to marketing campaigns. Furthermore, hyperparameter tuning was conducted using grid search to optimize parameters such as maximum depth and minimum sample split, ensuring the model was fine-tuned for the dataset used in this research.

3.5. Data Preprocessing

It is observed that the dataset is significantly imbalanced, with a considerably higher number of negative responses ("no") compared to positive responses ("yes"). This class imbalance poses a notable challenge because the model tends to predict the majority class more frequently. While this may lead to high overall accuracy, it results in poor identification of the minority class, which is crucial for the campaign's success [21].

To address the issue of class imbalance, a technique called resampling is implemented. Resampling involves adjusting the dataset to balance the class distribution, ensuring that the model has an equal representation of both classes during training. This can be achieved through various methods such as oversampling the minority class or undersampling the majority class [21]. In this study, the undersampling technique was specifically chosen due to its effectiveness in reducing the computational burden associated with large datasets. This approach involves randomly decreasing the number of instances in the majority class (negative responses) to match the number of instances in the minority class (positive responses). As a result, a more balanced dataset is created, allowing the model to learn the characteristics of both classes more effectively.

In addition to resampling, another effective approach that is used is adjusting the class weights [21]. By assigning higher weights to the minority class, the model enhances its sensitivity and recall towards positive responses, which is critical for correctly identifying successful marketing targets. This adjustment ensures that the model pays more attention to the minority class during the training process, thus improving its predictive capability.

To evaluate the effectiveness of these preprocessing techniques, performance metrics such as precision, recall, and F1-score were monitored during model evaluation. This provides a more comprehensive view of the model's performance, especially in its ability to identify the minority class, which is crucial for the marketing campaign's success.

3.6. Model Development

In the next part of the research, the DT model is developed using a structured and methodical approach. Initially, the dataset is prepared by partitioning the features into predictors (X) and the target variable (y). This method ensures that the model learns to predict the target variable based on the features [22]. The predictors consist of everything except the 'Response' column, which serves as the target variable. The dataset is divided into training and testing sets with an 80–20 ratio, meaning 80% of the data is used to train the model, and the remaining 20% is used to test it. This partitioning allows for evaluating the model's performance on unseen data, simulating real-world scenarios where the model will encounter new data. In this way, the model generalizes well and is not overfitted to the training data [22]. Additionally, a random state of 42 is specified to guarantee the

reproducibility of the results, ensuring that the random processes involved in data splitting will produce the same results every time the code is run.

Hyperparameter Tuning

After resampling, a grid search method, combined with cross-validation, is applied to explore different combinations of hyperparameters. One of the key ones is the ‘criterion’, which determines the function used to measure the quality of a split. The options for the ‘criterion’ parameter include Gini impurity and entropy [22].

Gini impurity is defined in Equation (1):

$$G = 1 - \sum_{i=1}^n p_i^2 \quad (1)$$

where p_i^2 represents the proportion of instances belonging to class i in the dataset. Gini impurity measures the probability of incorrectly classifying a randomly chosen element. An impurity of 0 indicates that all elements in a node belong to a single class, representing perfect purity. In practical terms, a lower Gini impurity means that the DT is better at creating homogeneous groups of customers, which can lead to more accurate predictions [22]. Entropy is defined in Equation (2):

$$H = - \sum_{i=1}^n p_i \log_2(p_i) \quad (2)$$

It measures the amount of disorder within a set of classes. When the entropy is 0, it means there is no disorder, and all customers within a node share the same classification. Higher entropy values indicate greater disorder and less purity. The criterion of entropy often leads to more balanced splits compared to Gini impurity, as it creates splits that increase the information gain, making it a preferred choice when the goal is to achieve higher accuracy and a more informative model [22].

Another important hyperparameter is the ‘splitter’. The ‘splitter’ can be set to ‘best’ or ‘random’. The ‘best’ option selects the optimal split among all features, aiming to maximize information gain or minimize Gini impurity. On the other hand, the ‘random’ option selects a random feature and then finds the best split within that feature. Parameter ‘best’ might result in a more accurate but computationally intensive model, whereas ‘random’ can lead to faster training times and increased generalization [22].

The ‘max_depth’ parameter controls the maximum depth of the tree. It ranges from no limit, allowing the tree to expand until all leaves are pure, to a specified maximum depth, such as 5, 10, 15, or 20. A shallower tree generalizes better on unseen data, whereas a deeper tree can capture more details from the training data but risks overfitting [22]. The ‘min_samples_split’ parameter specifies the minimum number of samples required to split an internal node. It ranges from 2 to 15. A higher value prevents the model from learning too much from the noise in the training data, thus improving its generalization capability [22].

Finally, the ‘min_samples_leaf’ parameter indicates the minimum number of samples required to be at a leaf node. It ranges from 1 to 6. A higher value can lead to a more generalized model, whereas a lower value might allow the tree to capture more patterns [22]. By conducting an exhaustive grid search across these parameters, the model is evaluated through cross-validation for each combination.

This means the model is trained and evaluated on different subsets of the training data to ensure that the hyperparameters are not overfitted to a particular subset. The cross-validation divides the training data into five parts, training the model in four parts and validating it on the fifth, rotating this process to cover all combinations [23].

The best combination of hyperparameters is identified based on the average performance across these folds [23]. The best estimator from the grid search is then selected as the final model (best_clf) for further evaluation.

3.7. Model Evaluation

Evaluating the performance of the predictive model is crucial for understanding how well it generalizes to new, unseen data. In this research, several key metrics are utilized to assess the effectiveness of the DT model in predicting customer responses to marketing campaigns. These metrics include accuracy, precision, recall, F1 score, and the confusion matrix.

3.7.1. Confusion Matrix

To gain a comprehensive understanding of a model's effectiveness in imbalanced scenarios, the use of a confusion matrix is essential. It summarizes the prediction results, showing the count of correct and incorrect predictions broken down by each class. True Positive (TP) refers to the number of instances where the model correctly predicts that a customer will respond positively to a campaign, aligning with actual positive responses. True Negative (TN) denotes cases where the model accurately identifies customers who will not respond, matching the actual negative responses. False Positives (FPs), often termed "false alarms" occur when the model incorrectly predicts a positive response from customers who, in reality, do not respond to the campaign. Conversely, False Negatives (FNs) happen when the model fails to predict a positive response from customers who indeed respond [24].

3.7.2. Accuracy

Accuracy is a measure of the overall correctness of the model [24], representing the proportion of correctly predicted instances out of the total instances, as shown in Equation (3).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

For this model, the accuracy indicates how well it can correctly classify both positive and negative responses. In the case of the imbalanced dataset in this study, high accuracy can be achieved by simply predicting the majority class most of the time. However, this high accuracy is deceptive because the model fails to identify the customers who respond, making it ineffective for practical purposes. The limitations of accuracy in the context of imbalanced datasets highlight the importance of alternative metrics such as precision, recall, and the F1 score [21].

3.7.3. Precision

Precision is also known as the positive predictive value. As defined in Equation (4), precision measures the accuracy of positive predictions [24].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

In this study, precision indicates the proportion of customers who are predicted to respond positively and indeed did respond positively.

3.7.4. Recall

As shown in Equation (5), recall measures the ability of the model to identify all actual positive instances [24].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

In this study, recall indicates the proportion of actual positive responses that were correctly predicted by the model.

3.7.5. F1-Score

The F1-score is the harmonic mean of precision and recall, providing a single metric that balances the two. It is particularly useful when there is an uneven class distribution [24]. The formula for the F1-score is shown in Equation (6):

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

The F1-score ranges from 0 to 1, where 1 indicates perfect precision and recall, and 0 indicates the worst possible performance. This metric is beneficial when seeking a balance between precision and recall, especially in the presence of class imbalance [24].

3.8. Feature Importance Extraction

In this step, feature importance scores are extracted from the trained DT model and the top 10 features are identified and visualized. Feature importance is a metric that indicates the significance of each input variable in contributing to the prediction accuracy of the DT classifier.

3.9. Decision Rules Generation

In this step, the decision tree rules are generated in the form of if.. else statements. They allow for easy interpretation of the decision-making process, where one can understand how a particular prediction is made. The clarity of the DT rules enables stakeholders, who may not have a deep technical background, not only to pinpoint these influential factors accurately but also to utilize them effectively.

4. Results

In this section of the study, the best hyperparameters resulting from the grid search combined with cross-validation are presented before and after resampling is applied. Additionally, a comparative analysis of the model evaluation results before and after resampling is conducted. The analysis focuses on the confusion matrix and various performance metrics, including accuracy, precision, recall, and F1-score, to evaluate the model's effectiveness. Furthermore, the results include feature importance scores and the generated decision rules, which are extracted from the decision tree classification model after resampling. This approach is taken because the resampled dataset provides a more balanced and accurate representation of the underlying patterns, leading to more reliable and interpretable decision rules and feature importance scores.

4.1. Best Hyperparameters

4.1.1. Before Resampling

The grid search combined with cross-validation identified the optimal hyperparameters to be the ones presented in Table 4 before resampling was applied.

Table 4. Parameter values before resampling.

Parameter	Value
criterion	entropy
max_depth	5
min_samples_leaf	2
min_samples_split	2
splitter	random

These hyperparameters reflect a conservative approach to handling the significant class imbalance in the dataset. The criterion of entropy helps in maximizing information gain at each split. By limiting the maximum depth to 5, the model avoids overfitting to the majority class of negative responses, which dominates the dataset. The parameters for minimum samples per leaf and split ensure that each node has enough data to make reliable decisions, thus reducing the likelihood of splits based on noise or anomalies. The use of a random splitter adds an element of randomness to the decision-making process, which helps prevent the model from becoming overly complex and biased towards the majority class during training.

4.1.2. After Resampling:

Following the application of undersampling to balance the class distribution, the grid search with cross-validation identified a different set of optimal hyperparameters, presented in Table 5.

Table 5. Parameter values after resampling.

Parameter	Value
criterion	entropy
max_depth	None
min_samples_leaf	2
min_samples_split	2
splitter	best

The shift in hyperparameters post-undersampling indicates a significant change in the model's complexity and its approach to decision-making. With the maximum depth set to none, the model is allowed to grow without constraints until all leaves are pure or until they contain fewer samples than the minimum samples split threshold. This unrestricted growth enables the model to capture more detailed patterns in the balanced dataset. The switch to the best splitter means the model now selects the optimal split at each node, based on the entropy criterion, to maximize information gain, leading to more precise and effective splits that better separate the classes.

4.2. Confusion Matrix

4.2.1. Before Resampling

The confusion matrix before resampling is presented in Table 6 and it reveals that the model correctly identifies 27 true positives and 357 true negatives, while there were 21 false positives and 36 false negatives.

Table 6. Confusion matrix before resampling.

Predicted \ Actual	Positive (+)	Negative (−)
Positive (+)	TP = 27	FP = 21
Negative (−)	FN = 36	TN = 357

This indicates that the model was successful in predicting customers who would respond positively to marketing campaigns in 27 instances and correctly identifying customers who would not respond in 357 instances. The high number of true negatives compared to true positives is attributed to the imbalance in the target class 'Response'. The model is exposed to more instances of non-response during training, which makes it better at identifying non-responders (true negatives) but limits its capacity to detect responders (true positives). The model also produced 21 False Positives (FPs), representing instances where the model incorrectly predicted a positive response from customers who did not respond. Conversely, the 36 False Negatives (FNs) indicate cases where the model failed to predict a positive response from customers who did respond positively. This means that the model occasionally mistakes non-responders for responders, potentially leading to unnecessary marketing efforts toward those unlikely to engage. More critically, the higher number of false negatives signifies that the model misses many potential customers who would have responded positively, ultimately resulting in missed opportunities for engagement.

4.2.2. After Resampling

The confusion matrix after resampling is presented in Table 7, and it reveals that the model correctly identifies 49 true positives and 51 true negatives, while there were 24 false positives and 10 false negatives.

Table 7. Confusion matrix after resampling.

Predicted \ Actual	Positive (+)	Negative (−)
Positive (+)	TP = 49	FP = 24
Negative (−)	FN = 10	TN = 51

Post-resampling, the model's ability to correctly identify positive responses improves significantly, evidenced by the increase in true positives from 27 to 49. This improvement is primarily due to the undersampling technique, which balances the class distribution by reducing the number of majority class instances, thereby allowing the model to learn more effectively from the minority class. However, while the model shows a notable reduction in false negatives (from 36 to 10), the increase in false positives (from 21 to 24) and the decrease in true negatives (from 357 to 51) indicate that some degree of imbalance remains. Specifically, the false positives represent almost half the true positives, demonstrating a trade-off typical in addressing class imbalance.

While the model becomes better at predicting positive responses, it sacrifices some accuracy in distinguishing non-responders, a common effect of undersampling. Despite this, the reduction in false negatives highlights that the model is better equipped to predict both responders and non-responders, though further refinements could help reduce false positives.

The breakdown in the confusion matrix is crucial for calculating performance metrics.

4.3. Model Evaluation

4.3.1. Before Resampling

The performance of the model before/after resampling is presented in Figure 5.

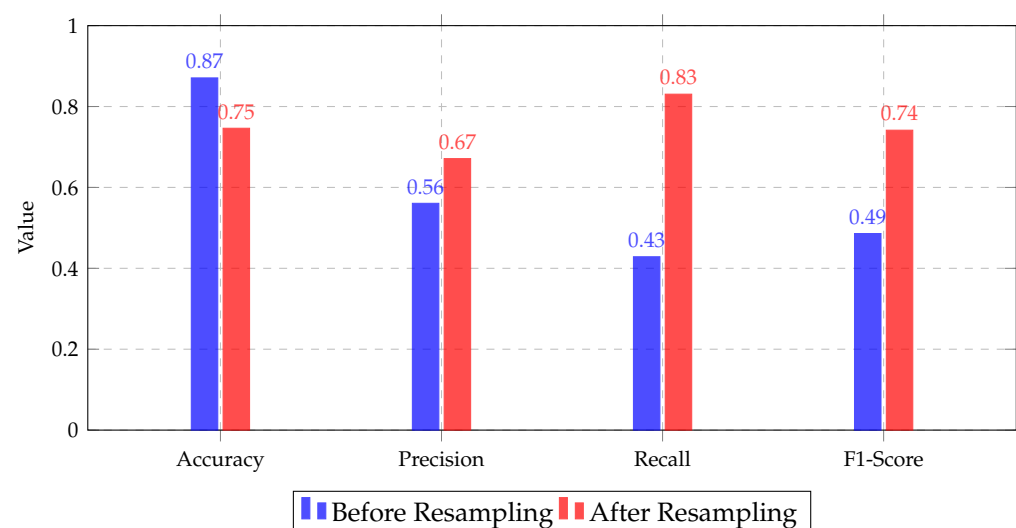


Figure 5. Comparison of evaluation metrics before and after resampling.

Despite the high accuracy of 87%, the precision, recall, and F1-score are relatively low. Accuracy alone can be misleading in cases of imbalanced datasets, where one class significantly outweighs the other. Here, the high accuracy mainly reflects the model's ability to correctly identify non-responders, but it does not adequately capture the performance in predicting the responders. The precision, which is calculated to be 56%, measures the

proportion of true positive predictions among all positive predictions. This means that out of all the instances that the model predicted as responders, only 56% were correct. The recall, calculated to be 44%, measures the proportion of actual positive instances that were correctly identified by the model. This means that the model only identified 44% of the actual responders correctly. The low F1-score reflects the overall inefficiency of the model in handling the imbalanced dataset, as it struggles to achieve a good trade-off between precision and recall. While the model appears to perform well based on accuracy alone, the low precision, recall, and F1-score reveal its limitations in predicting the minority class effectively.

4.3.2. After Resampling

The performance of the model after resampling is presented in Figure 5.

Post-resampling, the model's performance improved significantly. The accuracy dropped to 74.6%, which is expected, as the model now faces a more balanced dataset, making predictions more challenging. However, this decrease in accuracy is not necessarily a negative outcome. The balanced dataset has allowed for improvements in other critical metrics. The precision increased to 67.1%, indicating that the model is now better at correctly identifying true responders, reducing the number of false positives where non-responders are incorrectly predicted as responders. The recall increased to 83.1%, demonstrating a substantial improvement in capturing most of the true positive cases, thereby reducing the number of false negatives where actual responders are missed. Finally, the F1-score improved to 74.2%, providing a balanced measure of the model's precision and recall. The significant improvement in the evaluation metrics indicates that the model is now well-suited to identify both responders and non-responders accurately, making it more effective for practical applications in marketing campaigns.

4.4. Feature Importance Scores

The top 10 most influential features are presented in Figure 6. Demographic factors such as age and income are reported to play a crucial role in customer behavior. Past customer interactions with the company, indicated by variables like Recency (days since last purchase), Customer_Days (days since customer registration), and AcceptedCmpOverall (number of accepted campaigns), are significantly influential on customer response. Additionally, product-specific purchases such as MntGoldProds (spending on gold products) and MntMeatProducts (spending on meat products), along with purchase channels, including NumCatalogPurchases (number of catalog purchases), NumStorePurchases (number of store purchases), and NumWebPurchases (number of web purchases), influence the model's prediction of customer responses to direct marketing.

4.5. Decision Rules

A decision tree is a flowchart-like structure used for decision-making and predictive modeling. The tree consists of nodes that represent decisions or tests on features, branches that represent the outcomes of those tests, and leaf nodes that represent outcomes or classifications [25]. In this section, we elaborate on the decision rules generated by our model, specifically through the application of the decision tree algorithm. These rules illustrate how customer attributes influence the predicted outcomes. Understanding these rules not only enhances interpretability but also provides actionable insights for marketers. The rules follow a hierarchical structure where each decision is based on specific customer characteristics. The detailed decision tree rules are visualized in Appendix A, where they are divided into Algorithms A1–A5.

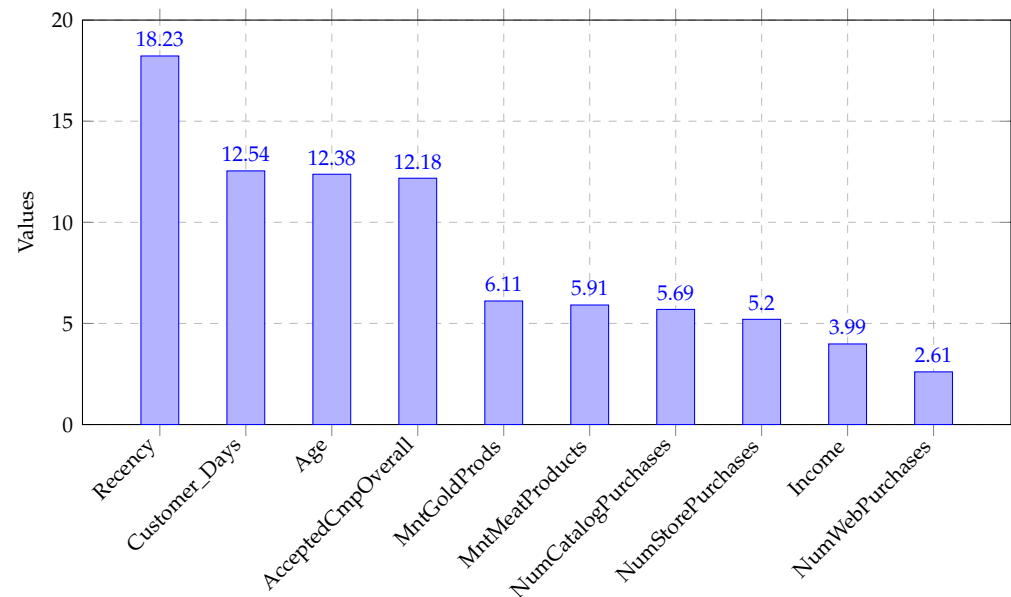


Figure 6. Feature importance scores.

4.6. Addressing the Research Questions

In this section, we revisit the Research Questions (RQs) outlined in the introduction and discuss how the results presented in the previous sections provide answers to these questions.

RQ1. What are the challenges and limitations presented in the literature regarding predicting customer marketing responses?

The challenges and limitations identified in the literature review are confirmed by the results obtained in this study. One of the main challenges is the class imbalance in the dataset, which was evident before applying the resampling technique. As shown in the confusion matrix before resampling (Table 6), the model struggled with a higher number of false negatives (36) and relatively low recall (44%). This highlights the difficulty of predicting positive customer responses when the majority of the data consists of non-responders. The results reinforce the need for techniques such as resampling to address such imbalances and improve model performance.

RQ2. How effective is the DT model at predicting customer response to marketing campaigns?

The effectiveness of the Decision Tree (DT) model in predicting customer responses is evaluated using performance metrics, such as accuracy, precision, recall, and F1-score, before and after resampling. Before resampling, the model exhibited high accuracy (87.3%), but precision, recall, and F1-score were relatively low (56%, 44%, and 49%, respectively), as shown in Figure 5. After resampling, the model's ability to predict positive responses improved significantly, with recall increasing to 83% and F1-score to 74%, despite a slight drop in accuracy to 75.3% (Figure 5). These improvements demonstrate that the DT model is effective, particularly after addressing the dataset imbalance through undersampling.

RQ3. What are the key factors influencing customer response to marketing campaigns as identified by the DT model?

The feature importance analysis (Figure 6) reveals the key factors influencing customer responses. Among the top 10 features, demographic factors such as age and income and past interactions like Recency (days since the last purchase) and AcceptedCmpOverall (number of accepted campaigns) are significant. This confirms that both customer characteristics

and their engagement history with the company play crucial roles in determining their likelihood to respond to marketing campaigns.

Which demographic factors are most influential in predicting customer response to marketing campaigns according to the DT model?

From the feature importance results, age and income are highlighted as the most influential demographic factors. This suggests that older customers with higher income levels may be more responsive to marketing efforts, aligning with findings in the existing literature on consumer behavior.

How do past interactions with the company affect future responses according to the DT model?

Past customer interactions, specifically Recency, Customer_Days, and AcceptedCampaignOverall, strongly influence future responses, as demonstrated by their high feature importance scores. Customers who have engaged more recently or have a history of accepting previous campaigns are more likely to respond positively to future marketing efforts, supporting the hypothesis that customer loyalty and previous engagement are key predictors of response.

5. Discussion

In this section, the results presented in Section 4 are interpreted, and their implications for marketing strategies are discussed.

5.1. Results Interpretation

The findings before resampling show an initial accuracy of 10.8%, indicating poor model performance in identifying positive responders. However, after resampling, precision increased to 67.0% and recall to 83.0%, directly supporting our conclusion that resampling significantly improves the model's predictive ability, enabling more effective targeting of potential responders. This is reflected in the relatively low precision (56%), recall (42.8%), and F1-score (48.6%), as well as in the confusion matrix that showed a significant number of false negatives (36) and a moderate number of false positives (21). The low accuracy primarily reflects the model's inability to correctly identify both responders and non-responders effectively. In this context, high accuracy is misleading, as it mainly reflects the model's ability to identify the majority class (non-responders), which does not align to predict the minority class of positive responders.

This imbalance necessitates the use of techniques to improve the model's sensitivity to the minority class. After applying resampling, the findings demonstrate a significant improvement in the model's ability to predict positive responses. The confusion matrix post-resampling shows a more balanced performance, with 49 true positives and 51 true negatives. Although the overall accuracy decreased to 54.4%, this decrease is expected and acceptable given the context of a more balanced dataset. The model's precision increased to 67.0%, indicating a higher proportion of correctly identified positive responders among all predicted positives. The recall improved dramatically to 83.0%, meaning the model is now much better at identifying actual responders, reducing the number of false negatives to 10. The F1-score also increased to 74.2%, providing a balanced measure of the model's precision and recall.

These improved results post-resampling mean that the model is now better suited to address the research questions related to predictive modeling in marketing campaigns. The substantial improvement in recall (from 42.8% to 83.0%) directly supports the conclusion that the model is now much more effective at identifying potential responders. This allows marketing teams to confidently allocate resources toward customers who are more likely to respond, minimizing unnecessary costs associated with targeting non-responders, maximizing the effectiveness of the campaigns and reducing unnecessary marketing expenses. The findings highlight the importance of balancing the dataset to improve model performance, ensuring that both responders and non-responders are effectively identified. Overall, the resampling approach has led to a more robust predictive model, capable of

providing actionable insights for marketing strategies. By focusing on the key influential features and understanding the dynamics of customer behavior, businesses can optimize their marketing efforts to achieve better engagement and conversion rates.

5.2. Implications for Marketing Strategies

In particular, the feature importance analysis in Figure 6 highlights several key factors influencing customer responses to marketing campaigns. Demographic factors, such as age and income, play significant roles. Age suggests that certain age groups are more likely to respond to marketing efforts. Income also impacts response rates, indicating that customers with higher income levels might engage more with marketing offers. Past interactions with the company are also really important in shaping the model's predictive power.

Recency is the most influential feature, suggesting that marketing efforts should focus on customers who have interacted with the company recently, as they are more likely to respond positively to new campaigns. Similarly, the duration of the customer's relationship with the company, measured by *Customer_Days*, indicates that long-term customers, who have developed loyalty, are more receptive to marketing initiatives. The acceptance of previous campaigns (*AcceptedCmpOverall*) reflects customers' historical engagement with marketing efforts, suggesting that those who have positively responded in the past are more likely to do so in the future. Additionally, specific product categories, such as *MntGoldProds* and *MntMeatProducts*, influence customer responses, indicating preferences for certain products.

Understanding these preferences allows for more effective product-specific promotions. The results in this study align with the findings of previous studies, such as those by Apampa [14] and Choi et al. [10], which also highlighted the importance of demographic and past interaction data. However, our study found that Recency and *Customer_Days* were more influential than previously reported, possibly due to the specific characteristics of our dataset and the context of the marketing campaigns analyzed. Furthermore, the model is interpretable, providing clear and understandable decision rules. This interpretability is a significant advantage in the context of marketing campaigns. For example, one of the key decision rules, visualized in Appendix A, particularly in Algorithm A1, indicates that if a customer has accepted half of the previous campaigns ($\text{AcceptedCmpOverall} \leq 0.50$), the model then considers their recency of interaction ($\text{Recency} \leq 42.50$). If the customer has interacted with the company in the past 42 days, the model further refines its decision based on the number of catalog purchases ($\text{NumCatalogPurchases} \leq 0.50$). Such rules are straightforward and easily comprehensible for marketing professionals, enabling them to understand the logic behind the model's predictions and make informed decisions based on these insights.

This clarity builds trust in the model's recommendations. Marketing teams can confidently use the model to target customers, knowing that the predictions are based on logical and understandable criteria. This transparency is crucial for the practical application of the predictive models. Moreover, the interpretability ensures that the model can be easily updated and adjusted as new data become available. As marketing campaigns evolve and customer behaviors change, the decision rules can be re-evaluated and refined.

5.3. Limitations

Despite the improvements achieved through resampling and the valuable insights provided by the model, there are several limitations to consider:

- **Limited Dataset:** While the results demonstrate strong model performance within the dataset, the focus on food companies limits the generalizability of these findings. Further evaluation of diverse datasets is necessary to fully assess the model's applicability to other industries, as highlighted in the proposed future work. The unique characteristics of this dataset may not accurately represent customer behavior across different industries or sectors, warranting caution when applying the model's conclusions beyond the food industry.

- **Computational Efficiency:** There is a significant need to detail the computational efficiency of the models, particularly in real-world scenarios. As noted in the paper, the trade-off between computational demands and accuracy gains is crucial, especially with gradient boosting compared to simpler models like decision trees and random forests. The lack of detailed information on training times and resource utilization could hinder practical implementation considerations.

5.4. Comparison

In this subsection, we compare our proposed solution with the approaches and models discussed in the related works. This comparison aims to highlight the advancements, advantages, and unique contributions of our solution in the context of predictive models for direct marketing.

5.4.1. Overview of Related Works

Table 8 offers a detailed comparison of our proposed solution with various predictive models presented in the literature. Our gradient boosting model achieved the highest accuracy of 91.5%, outperforming decision tree and random forest models. This high performance is particularly noteworthy in handling imbalanced datasets, where many of the related works faced challenges. The table also highlights the trade-offs between accuracy, model complexity, and interpretability across different studies. A complete comparison and in-depth analysis of these results can be found in Section 5.4.2.

Table 8. Benchmark of related work with our proposed solution.

Study	Models Compared	Best Accuracy Achieved	Key Findings
K. Wisaeng [15]	SVM, J48-graft, LAD tree, RBFN	86.95% (SVM)	SVM outperformed other models, RBFN had the lowest accuracy.
Sérgio Moro et al. [16]	Logistic regression, neural networks, decision tree, SVM	79% (Neural networks)	Neural networks performed best in predicting customer behavior; targeting top half of customers improved outcomes.
Sérgio Moro [9]	Naive Bayes, decision tree, SVM	N/A (SVM)	SVM showed the highest prediction performance, with call duration as the most significant feature.
Choi et al. [10]	Decision tree	87.23% (Non-responders)	Decision tree accurately predicted non-responders but had lower accuracy for predicting responders.
Usman-Hamza et al. [18]	Decision tree, naive Bayes, K-nearest neighbor	93.6% (Decision tree)	Decision tree outperformed other models in customer churn analysis for live stream e-commerce.
Chaubey et al. [1]	Random forest, decision tree	N/A (Random forest)	Random forest performed better than decision tree for churn prediction but lacks interpretability.

Table 8. Cont.

Study	Models Compared	Best Accuracy Achieved	Key Findings
Olson et al. [17]	RFM, logistic regression, decision tree, neural networks	N/A (Decision tree)	Advanced models like decision tree and neural network outperformed the traditional RFM model. Decision trees offered the best balance between accuracy and interpretability.
Liu & Yang [11]	Decision tree, SVM, neural network, naive Bayes, random forest	91.35% (Decision tree)	Decision tree achieved the highest accuracy in content marketing prediction; precision and F1 score also favored the decision tree model.
Safarkhani & Moro [12]	Decision tree, naive Bayes, logistic regression	94.39% (Decision tree)	Decision tree outperformed other models, achieving the highest accuracy in predicting customer behavior for bank telemarketing.
Asare-Frempong & Jayabalan [13]	Random forest, decision tree, logistic regression, MLPNN	86.8% (Random forest)	Random forest outperformed other models, while decision tree provided useful interpretability for customer term deposit subscription prediction. Random forest did not consistently improve decision tree's performance in bank marketing prediction, favoring decision tree for interpretability.
Apampa [14]	Random forest, decision tree	N/A (Decision tree)	
Our proposed Solution	Decision tree, random forest, gradient boosting	91.5% (Gradient boosting)	Gradient boosting achieved the highest accuracy, outperforming decision tree and random forest; effective in handling imbalanced datasets.

5.4.2. Comparison with Our Proposed Solution

As highlighted in Table 8, our proposed gradient boosting model stands out for its superior performance in terms of accuracy (91.5%). This performance, compared with existing works, indicates its strength in handling the challenges of imbalanced datasets, a common issue in direct marketing predictions. For instance, Usman-Hamza et al. [18] achieved 93.6% with decision tree, but their model does not explicitly address imbalanced datasets as effectively as ours. Furthermore, while some models like neural networks (e.g., Sérgio Moro et al. [16]) excelled in accuracy, their interpretability is limited compared to decision tree and random forest, which are more transparent yet less accurate.

The main findings from the table indicate that while different models have varying strengths (e.g., high accuracy in specific scenarios), our approach with gradient boosting provides a well-rounded solution that balances accuracy and the ability to manage imbalanced data. However, it is essential to note that this comes at the cost of higher computational complexity, a factor also discussed in previous studies (e.g., K. Wisaeng [15] and Apampa [15]), where simpler models like decision tree are preferred for their interpretability and efficiency.

5.4.3. Summary and Implications

Our proposed solution demonstrates a significant improvement in predictive accuracy compared to many of the related works, particularly through the use of gradient boosting. This advancement highlights the potential of leveraging more sophisticated models for direct marketing predictions while balancing the trade-offs between accuracy, interpretability, and computational efficiency. The results underscore the effectiveness of our approach to enhancing predictive performance and provide valuable insights into the ongoing evolution of predictive modeling in this domain. For a detailed comparison of related works, refer to Table 8.

6. Conclusions

This study demonstrates the effectiveness of using DT models for predicting customer responses to marketing campaigns. By addressing the challenges of class imbalance through resampling and adjusting class weights, the model's ability to accurately predict positive responses improved significantly. This research not only identifies key demographic and interaction factors influencing customer behavior but also provides a transparent and interpretable model, crucial for practical applications in marketing strategies.

The study answers the three primary research questions as follows:

1. The first question regarding the challenges and limitations presented in the literature is addressed in Section 2, highlighting the complexities of customer behavior and the limitations of traditional predictive models.
2. The second question on the effectiveness of the DT model in predicting customer response to marketing campaigns is explored through the comparative analysis of model evaluation metrics before and after resampling, as presented in Sections 4.2 and 4.3, and interpreted in Section 5.1. This includes a detailed examination of how resampling techniques improved the model's ability to predict positive responses.
3. The key factors influencing customer response are identified through feature importance analysis and decision rules extraction, presented in Section 4.4 and discussed in Section 5.2. This analysis provides insights into the most significant demographic and interaction factors affecting customer responses.

Despite the significant improvements achieved, there are several limitations to this study. The dataset, while comprehensive, is limited to a specific context and may not generalize to other industries or geographical regions. In future research, we plan to test the model with datasets from diverse sectors and regions to assess its generalizability beyond the iFood platform and ensure broader applicability to different industries. We are also planning to evaluate the computational efficiency of our model, focusing on training times, memory usage, and scalability across various datasets. This will provide a more thorough understanding of the model's practical applicability in real-world scenarios. Additionally, the use of undersampling, while effective in balancing the classes, reduces the overall dataset size, potentially excluding valuable information from the majority class. Future research should explore the integration of ensemble methods to improve model performance, as studies have shown that ensemble methods, such as RF, can provide significant improvements in handling imbalanced datasets and improving prediction accuracy [26].

Author Contributions: Conceptualization, M.E.-H. and M.P.; methodology, M.E.-H.; software, M.P.; validation, M.E.-H. and M.P.; formal analysis, M.E.-H. and M.P.; investigation, M.E.-H. and M.P.; resources, M.P.; data curation, M.P.; writing—original draft preparation, M.E.-H. and M.P.; writing—review and editing, M.E.-H. and M.P.; visualization, M.P.; supervision, M.E.-H.; project administration, M.E.-H.; funding acquisition, M.E.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available in the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DT	Decision Tree
CRM	Customer Relationship Management
ML	Machine Learning
RF	Random Forest
RBFN	Radial Basis Function Network
SVM	Support Vector Machine
LR	Logistic Regression
NN	Neural Network
NB	Naive Bayes
MLPNN	Multilayer Perceptron Neural Network
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative

Appendix A

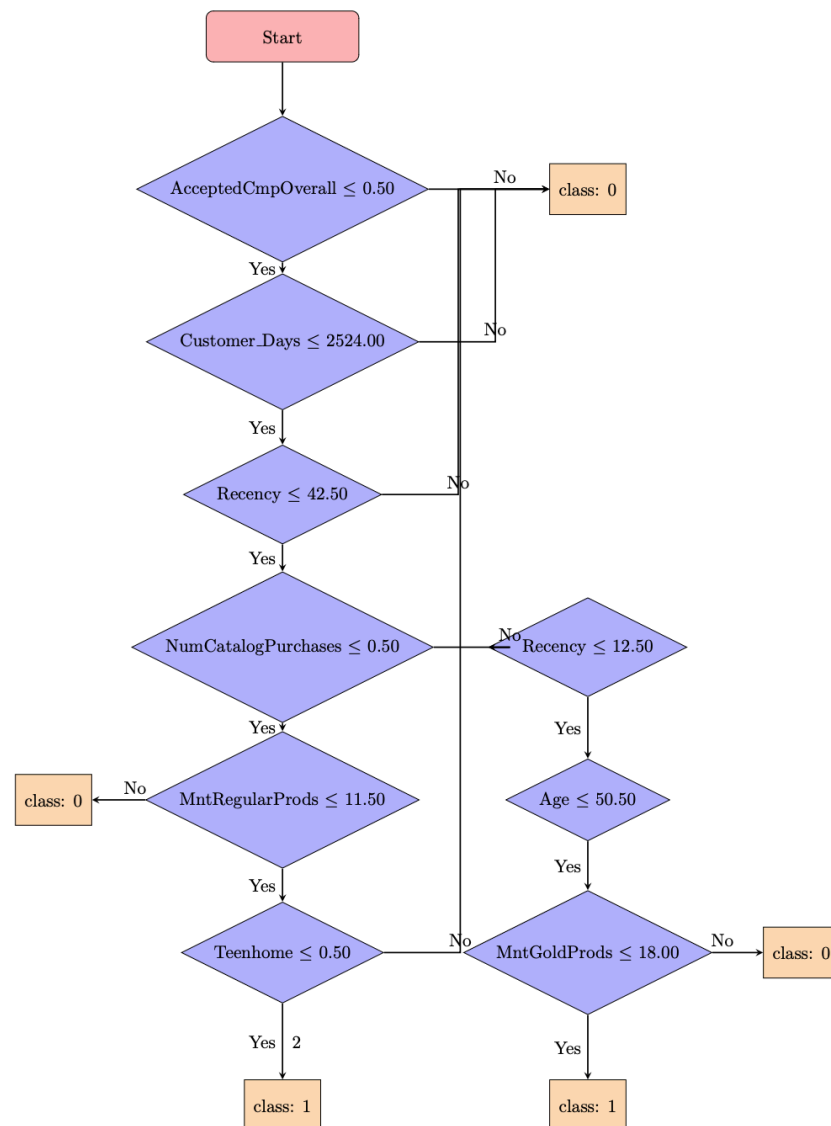


Figure A1. Flowchart representing a decision tree algorithm.

Algorithm A1 Decision Tree Rules Part 1

```

1: if AcceptedCmpOverall  $\leq$  0.50 then
2:   if Customer_Days  $\leq$  2524.00 then
3:     if Recency  $\leq$  42.50 then
4:       if NumCatalogPurchases  $\leq$  0.50 then
5:         if MntRegularProds  $\leq$  11.50 then
6:           if Teenhome  $\leq$  0.50 then
7:             class: 1
8:           else
9:             class: 0
10:          end if
11:        else
12:          class: 0
13:        end if
14:      else
15:        if Recency  $\leq$  12.50 then
16:          if Age  $\leq$  50.50 then
17:            if MntGoldProds  $\leq$  18.00 then
18:              if Customer_Days  $\leq$  2411.00 then
19:                class: 1
20:              else
21:                class: 1
22:              end if
23:            else
24:              class: 1
25:            end if
26:          else
27:            if Age  $\leq$  60.50 then
28:              class: 0
29:            else
30:              if Customer_Days  $\leq$  2245.50 then
31:                class: 1
32:              else
33:                class: 1
34:              end if
35:            end if
36:          end if
37:        else
38:          if marital_Together  $\leq$  0.50 then
39:            if MntSweetProducts  $\leq$  68.00 then
40:              if Teenhome  $\leq$  0.50 then
41:                if Age  $\leq$  48.00 then
42:                  if Age  $\leq$  41.50 then
43:                    if MntWines  $\leq$  167.50 then
44:                      ... (truncated branch of depth 2)
45:                    else
46:                      class: 1
47:                    end if
48:                  else
49:                    class: 0
50:                  end if
51:                else
52:                  class: 1
53:                end if
54:              else
55:                if Age  $\leq$  52.00 then
56:                  class: 1
57:                else
58:                  class: 0
59:                end if
60:              end if
61:            else
62:              class: 0
63:            end if
64:          else
65:            class: 0
66:          end if
67:        end if
68:      end if
69:    else
70:      if MntMeatProducts  $\leq$  661.00 then
71:        class: 0
72:      else

```

Algorithm A1 Cont.

```

73:         class: 1
74:     end if
75: end if
76: end if
77: end if

```

Algorithm A2 Decision Tree Rules Part 2

```

1: if AcceptedCmpOverall  $\leq$  0.50 then
2:   if Customer_Days  $\leq$  2524.00 then ...
3:   else
4:     if Recency  $\leq$  36.50 then
5:       if Age  $\leq$  32.50 then
6:         class: 0
7:       else
8:         if MntMeatProducts  $\leq$  7.50 then
9:           class: 0
10:        else
11:          if Age  $\leq$  74.00 then
12:            if Income  $\leq$  48,828.00 then
13:              if NumDealsPurchases  $\leq$  2.50 then
14:                if Customer_Days  $\leq$  2712.00 then
15:                  if Income  $\leq$  21,170.00 then
16:                    class: 1
17:                  else
18:                    class: 1
19:                  end if
20:                else
21:                  class: 0
22:                end if
23:              else
24:                class: 1
25:              end if
26:            else
27:              if MntMeatProducts  $\leq$  498.00 then
28:                if Age  $\leq$  65.50 then
29:                  if NumStorePurchases  $\leq$  6.50 then
30:                    class: 0
31:                  else
32:                    if MntGoldProds  $\leq$  175.50 then
33:                      ... (truncated branch of depth 2)
34:                    else
35:                      class: 0
36:                    end if
37:                  end if
38:                else
39:                  class: 0
40:                end if
41:              else
42:                if education_Master  $\leq$  0.50 then
43:                  class: 1
44:                else
45:                  class: 1
46:                end if
47:              end if
48:            end if
49:          else
50:            class: 0
51:          end if
52:        end if
53:      end if
54:    end if
55:  end if
56: end if

```

Algorithm A3 Decision Tree Rules Part 3

```

1: if AcceptedCmpOverall  $\leq$  0.50 then
2:   if Customer_Days  $\leq$  2524.00 then ...
3:   else
4:     if Recency  $\leq$  36.50 then ...
5:     else
6:       if Customer_Days  $\leq$  2744.00 then
7:         if Income  $\leq$  78,112.50 then
8:           if Age  $\leq$  45.50 then
9:             class: 0
10:          else
11:            if MntFishProducts  $\leq$  5.00 then
12:              if Recency  $\leq$  45.00 then
13:                class: 0
14:              else
15:                if Age  $\leq$  69.00 then
16:                  if MntFruits  $\leq$  2.50 then
17:                    class: 1
18:                  else
19:                    class: 1
20:                  end if
21:                else
22:                  class: 0
23:                end if
24:              end if
25:            else
26:              if MntGoldProds  $\leq$  107.50 then
27:                class: 0
28:              else
29:                if MntFruits  $\leq$  25.50 then
30:                  class: 0
31:                else
32:                  class: 1
33:                end if
34:              end if
35:            end if
36:          end if
37:        else
38:          class: 1
39:        end if
40:      else
41:        if Age  $\leq$  61.00 then
42:          if NumWebPurchases  $\leq$  9.50 then
43:            if NumWebPurchases  $\leq$  5.50 then
44:              if MntGoldProds  $\leq$  13.00 then
45:                class: 0
46:              else
47:                if MntGoldProds  $\leq$  76.50 then
48:                  if Age  $\leq$  35.50 then
49:                    class: 0
50:                  else
51:                    if marital_Single  $\leq$  0.50 then
52:                      ... (truncated branch of depth 3)
53:                    else
54:                      ... (truncated branch of depth 2)
55:                    end if
56:                  end if
57:                else
58:                  class: 0
59:                end if
60:              end if
61:            else
62:              class: 1
63:            end if
64:          else
65:            class: 0
66:          end if
67:        else
68:          class: 1
69:        end if

```

Algorithm A3 Cont.

```

70:         end if
71:     end if
72: end if
73: end if

```

Algorithm A4 Decision Tree Rules Part 4

```

1: if AcceptedCmpOverall  $\leq$  0.50 then ...
2: else
3:     if Recency  $\leq$  22.50 then
4:         if Teenhome  $\leq$  0.50 then
5:             class: 1
6:         else
7:             if Income  $\leq$  66,106.50 then
8:                 class: 1
9:             else
10:                if AcceptedCmp4  $\leq$  0.50 then
11:                    class: 1
12:                else
13:                    class: 0
14:                end if
15:            end if
16:        end if
17:    else
18:        if Customer_Days  $\leq$  2421.00 then
19:            if Recency  $\leq$  75.00 then
20:                if Recency  $\leq$  24.50 then
21:                    class: 0
22:                else
23:                    if MntTotal  $\leq$  1367.00 then
24:                        if NumStorePurchases  $\leq$  3.50 then
25:                            class: 1
26:                        else
27:                            if MntWines  $\leq$  595.00 then
28:                                class: 0
29:                            else
30:                                class: 0
31:                            end if
32:                        end if
33:                    else
34:                        if NumStorePurchases  $\leq$  11.50 then
35:                            class: 1
36:                        else
37:                            class: 1
38:                        end if
39:                    end if
40:                end if
41:            else
42:                class: 0
43:            end if
44:        end if
45:    end if
46: end if

```

Algorithm A5 Decision Tree Rules Part 5

```

1: if AcceptedCmpOverall  $\leq$  0.50 then ...
2: else
3:   if Recency  $\leq$  22.50 then ...
4:   else
5:     if Customer_Days  $\leq$  2421.00 then ...
6:     else
7:       if NumCatalogPurchases  $\leq$  7.50 then
8:         if education_Graduation  $\leq$  0.50 then
9:           if MntTotal  $\leq$  216.00 then
10:            if MntFruits  $\leq$  1.50 then
11:              class: 0
12:            else
13:              class: 1
14:            end if
15:          else
16:            if MntMeatProducts  $\leq$  467.00 then
17:              if MntMeatProducts  $\leq$  76.00 then
18:                class: 1
19:              else
20:                if MntTotal  $\leq$  897.50 then
21:                  if AcceptedCmp5  $\leq$  0.50 then
22:                    class: 0
23:                  else
24:                    class: 1
25:                  end if
26:                else if
27:                  if Age  $\leq$  64.50 then
28:                    class: 1
29:                  else
30:                    if NumStorePurchases  $\leq$  11.50 then
31:                      class: 0
32:                    else
33:                      class: 1
34:                    end if
35:                  end if
36:                end if
37:              end if
38:            else
39:              class: 1
40:            end if
41:          end if
42:        else
43:          if Income  $\leq$  31172.00 then
44:            class: 1
45:          else
46:            if NumWebPurchases  $\leq$  2.50 then
47:              class: 0
48:            else
49:              if Age  $\leq$  53.00 then
50:                if AcceptedCmpOverall  $\leq$  1.50 then
51:                  if NumCatalogPurchases  $\leq$  3.50 then
52:                    if NumStorePurchases  $\leq$  7.50 then
53:                      class: 1
54:                    else
55:                      class: 0
56:                    end if
57:                  else
58:                    class: 0
59:                  end if
60:                else
61:                  class: 1
62:                end if
63:              else
64:                class: 0
65:              end if
66:            end if
67:          end if
68:        end if
69:      else

```

Algorithm A5 Cont.

```

70:         if Teenhome  $\leq$  0.50 then
71:             class: 1
72:         else
73:             class: 1
74:         end if
75:     end if
76: end if
77: end if
78: end if

```

References

- Chaubey, G.; Gavhane, P.R.; Bisen, D.; Arjaria, S.K. Customer purchasing behavior prediction using machine learning classification techniques. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 16133–16157. [\[CrossRef\]](#)
- Al Khaldy, M.A.; Al-Obaydi, B.A.A.; al Shari, A.J. The Impact of Predictive Analytics and AI on Digital Marketing Strategy and ROI. In *The Palgrave Handbook of Interactive Marketing*; Springer: Cham, Switzerland, 2023. [\[CrossRef\]](#)
- Song, Y.-y.; Lu, Y. Decision tree methods: Applications for classification and prediction. *Shanghai Arch Psychiatry* **2015**, *27*, 130–135. [\[CrossRef\]](#) [\[PubMed\]](#)
- Raorane, A.; Kulkarni, R. Data mining techniques: A source for consumer behavior analysis. *arXiv* **2011**, arXiv:1109.1202. [\[CrossRef\]](#)
- Louppe, G. Understanding Random Forests: From Theory to Practice. Ph.D. Thesis, University of Liège, Liège, Belgium, 2014.
- Kursa, M.B.; Rudnicki, W.R. The All Relevant Feature Selection using Random Forest. *arXiv* **2011**, arXiv:1106.5112. [\[CrossRef\]](#)
- Michal Moshkovitz, Y.Y.Y.; Chaudhuri, K. Connecting Interpretability and Robustness in Decision Trees through Separation. *arXiv* **2021**, arXiv:2102.07048. [\[CrossRef\]](#)
- Reinartz, W.J.; Kumar, V. The mismanagement of customer loyalty. *Harv. Bus. Rev.* **2002**, *80*, 86–94. [\[PubMed\]](#)
- Sérgio Moro, R.M.S.L.; Cortez, P. *Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology*; Technical Report; Universidade do Minho: Lisboa, Portugal, 2011.
- Youngkeun Choi, S.; Choi, J.W. Assessing the Predictive Performance of Machine Learning in Direct Marketing Response. *Int. J. E-Bus. Res.* **2023**, *19*, 1–12. [\[CrossRef\]](#)
- Liu, Y.; Yang, S. Application of Decision Tree-Based Classification Algorithm on Content Marketing. *J. Math.* **2022**, *2022*, 1–10. [\[CrossRef\]](#)
- Safarkhani, F.; Moro, S. Improving the Accuracy of Predicting Bank Depositor's Behavior Using a Decision Tree. *Appl. Sci.* **2021**, *11*, 9016. [\[CrossRef\]](#)
- Asare-Frempong, J.; Jayabalan, M. Predicting Customer Response to Bank Direct Telemarketing Campaign. In Proceedings of the 2017 International Conference on Engineering Technology and Technopreneurship (ICE2T), Kuala Lumpur, Malaysia, 18–20 September 2017.
- Apampa, O. Evaluation of Classification and Ensemble Algorithms for Bank Customer Marketing Response Prediction. *J. Int. Technol. Inf. Manag.* **2016**, *25*, 6. (In English) [\[CrossRef\]](#)
- Wisaeng, K. A Comparison of Different Classification Techniques for Bank Direct Marketing. *Int. J. Soft Comput. Eng. (IJSCE)* **2013**, *3*, 116–119.
- Sérgio Moro, P.C.; Rita, P. A data-driven approach to predict the success of bank telemarketing. *Int. J. Soft Comput. Eng. (IJSCE)* **2014**, *62*, 22–31.
- Olson, D.L.; Chae, B.K. Direct Marketing Decision Support through Predictive Customer Response Modeling. *J. Decis. Support Syst.* **2012**, *54*, 443–451. [\[CrossRef\]](#)
- Usman-Hamza, F.E.; Balogun, A.O.; Nasiru, S.K.; Capretz, L.F.; Mojeed, H.A.; Salihu, S.A.; Akintola, A.G.; Mabayoje, M.A.; Awotunde, J.B. Empirical analysis of tree-based classification models for customer churn prediction. *Sci. Afr.* **2024**, *23*, e02054. [\[CrossRef\]](#)
- iFood. iFood DF. 2024. Available online: <https://www.kaggle.com/datasets/diniwilliams/ifood-df> (accessed on 10 March 2024).
- iFood Restaurants Data—kaggle.com. Available online: <https://www.kaggle.com/datasets/ricardotachinardi/ifood-restaurants-data> (accessed on 2 March 2024).
- He, H.; Garcia, E.A. Learning from imbalanced data: Open challenges and future directions. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
- Fürnkranz, J. Decision Tree. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2011; pp. 263–267. [\[CrossRef\]](#)
- Refaeilzadeh, P.; Tang, L.; Liu, H. Cross-Validation. In *Encyclopedia of Database Systems*; Liu, L., Özsu, M.T., Eds.; Springer: Boston, MA, USA, 2009; pp. 532–538. [\[CrossRef\]](#)
- Hossin, M.; Sulaiman, M.N. A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1–11.

25. Suthaharan, S.; Suthaharan, S. Decision tree learning. *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Springer: Boston, MA, USA, 2016; pp. 237–269.
26. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.