UDC 681
DOI: 10.15587/1729-4061.2025.327158

OPTIMIZATION OF MARKETING CAMPAIGNS USING A MODIFIED ID3 DECISION TREE ALGORITHM

Asrianda
Corresponding author
Doctoral Student Program in Computer Science*
E-mail: asrianda@students.usu.ac.id
Herman Mawengkang
Professor*
Poltak Sihombing
Professor*
Mahyuddin K. M. Nasution
Professor*
*Department of Computer Science
Universitas Sumatera Utara
Dr. T. Mansur str., 9, Sumatera Utara, Indonesia, 20155

The object of this study is the strategy of online retail marketing campaigns, particularly in the context of utilizing a modified ID3 decision tree algorithm to improve predictive effectiveness regarding consumer responses. It addresses challenges in audience segmentation, campaign evaluation, and market adaptation, while also tackling technical issues such as overfitting, prediction errors, and data imbalance. These challenges often hinder businesses from accurately identifying and targeting potential customers, leading to inefficient marketing strategies and resource allocation. The dataset was split into 80:20 and 70:30 ratios, and the model was tested across decision tree depths from max_depth 1 to max_depth 20. The highest accuracy occurred at max_depth 6, ensuring optimal computational efficiency. However, increasing tree depth led to declining accuracy and rising computational costs, highlighting the risk of overfitting. Key factors influencing consumer response include income, education level, and recent company interactions. These variables help determine purchasing behavior and engagement levels, making them crucial in refining marketing strategies. Class imbalance introduces bias, affecting model performance by favoring the majority class while underrepresenting minority groups. The modified ID3 model outperforms ID3 Shannon, offering better precision for the majority class but lower recall for the minority class. Limiting campaign offers to one or two improves consumer responsiveness and prevents information overload. A data-driven marketing strategy ensures promotions align with consumer preferences and market trends. The developed model enables businesses to better target campaigns, increase conversion rates, and optimize resource allocation, ensuring an effective balance between tree depth and model accuracy

***Keywords:** marketing campaign, ID3, decision tree, class imbalance, accuracy, entropy modification, overfitting, data splitting, majority class, minority class*

Received 28.01.2025
Received in revised form 12.03.2025
Accepted 11.04.2025
Published 22.04.2025

***How to Cite:** Asrianda, Mawengkang, H., Sihombing, P., Nasution, M. K. M. (2025). Optimization of marketing campaigns using a modified ID3 decision tree algorithm. Eastern-European Journal of Enterprise Technologies, 2 (13 (134)), 58–70. https://doi.org/10.15587/1729-4061.2025.327158*

1. Introduction

In the current era of digital transformation, marketing campaigns play a vital role in shaping consumer perceptions and guiding purchasing decisions. As online retail continues to expand, companies must develop more targeted and efficient strategies to maintain relevance and competitiveness. However, many marketing campaigns still face critical issues, such as the inability to identify the right target audience, resulting in misdirected messages and reduced effectiveness. These problems are further exacerbated by poor segmentation, generic messaging, and inefficient allocation of marketing resources, ultimately leading to suboptimal outcomes and financial losses. Moreover, regulatory risks and misalignment between marketing strategies and dynamic market conditions can significantly harm brand reputation and performance [1–3].

Given these challenges, it is increasingly important to explore data-driven methods that can enhance campaign accuracy and impact. These studies are particularly relevant in modern conditions, where consumers are exposed to an overwhelming volume of digital content and demand more personalized and meaningful communication. While much of the existing literature has emphasized the success factors of marketing campaigns [4], limited attention has been paid to understanding failure points or exploring new algorithmic approaches to improve marketing interventions.

The advancement of online retail marketing has transformed how consumers process information, requiring practitioners and policymakers to understand the structural elements that shape campaign effectiveness [5, 6]. Key components such as campaign framing, source credibility, and segmentation significantly influence consumer behavior but are often overlooked [7–9]. This gap between theoretical frameworks and real-world application necessitates research that connects marketing concepts to measurable outcomes.

In this context, decision tree algorithms offer practical utility by enabling predictive modeling based on customer attributes. They can help marketers predict campaign success and optimize targeting strategies [10, 11]. However, challenges such as overfitting, poor representation of minority classes, and the limitations of existing algorithms like ID3 remain unresolved [12–14].

Addressing challenges in developing models for marketing campaigns requires an approach that is not only accurate but also adaptive to changing market dynamics. The primary obstacle lies in identifying the right audience, as misaligned segmentation strategies can render campaign messages ineffective. Challenges in data processing, such as overfitting and class imbalance within datasets, can reduce the reliability of predictive models. Additionally, shifting consumer trends, marketing regulations, and economic uncertainty further complicate the formulation of effective marketing strategies. Modifying the entropy function in the ID3 decision tree algorithm is necessary to enhance accuracy, improve predictions, and reduce bias. By overcoming these challenges, the developed model can effectively assist businesses in adjusting their marketing strategies and optimizing resource allocation efficiently. The studies on the development of marketing campaigns using the ID3 decision tree algorithm with entropy modification is relevant, for enhancing the effectiveness of marketing strategies in addressing dynamic market challenges.

2. Literature review and problem statement

Previous studies have examined the effectiveness of marketing campaigns in enhancing consumer engagement and strengthening brand loyalty [15]. These studies highlight the utilization of real-time data, enabling companies to dynamically adjust their marketing strategies. Another study [16] emphasizes the importance of understanding consumer behavior, including purchasing habits and responses to communication channels, underscoring how the evolution of marketing trends provides insights into consumer preferences and helps companies tailor their messaging effectively.

Several unresolved challenges remain in the implementation of marketing strategies. One of these challenges is the inefficiency in selecting the right target audience. Previous studies [17] indicate that targeting errors lead to resource wastage and a decline in campaign effectiveness. Weak or inconsistent branding has also been shown to reduce campaign appeal. Another challenge is the difficulty in objectively measuring the effectiveness of marketing strategies. External factors such as economic dynamics and shifts in social trends often influence campaign outcomes. Companies must conduct continuous market analysis to adapt their marketing strategies to evolving conditions.

These limitations may arise from several factors. First, the lack of integration of accurate analytical methods in identifying consumer behavior patterns. Data processing techniques used in previous studies have failed to capture the complexity of interactions between various marketing variables. Second, the challenge of handling imbalanced datasets in digital marketing, which results in models favoring majority consumer groups over niche market segments that, while smaller, hold high value. Previous research [18] reveals that the ID3 algorithm, commonly used in marketing decision-making, exhibits bias toward attributes with multiple values, leading to less accurate models. This bias results in the selection of less informative attributes, reducing the model's effectiveness in classifying data optimally.

This study proposes a novel approach by modifying the ID3 algorithm with optimized entropy. The model is designed to handle imbalanced datasets more accurately and improve interpretability in consumer pattern evaluation. To address computational efficiency limitations, an entropy calculation method is required to reduce the complexity of processing large-scale data.

Several previous studies have attempted to overcome these challenges using various methods. Studies [19, 20] highlight the issue of multi-value bias in attribute selection, demonstrating that overly complex models tend to suffer from overfitting. There have been prior efforts to enhance the effectiveness of ID3 in digital marketing. However, existing approaches have yet to be optimized for handling large and imbalanced datasets. Further exploration is needed to develop a model that is both adaptive and efficient in processing marketing data.

Based on these identified gaps, this study aims to bridge the shortcomings of previous research by developing an improved ID3 algorithm capable of handling complex datasets effectively. With a more accurate approach to evaluating consumer patterns, this study is expected to make a significant contribution to enhancing the effectiveness of marketing campaigns and overall business strategies.

3. The aim and objectives of the study

The aim of the study is to enhance the effectiveness of marketing campaigns by utilizing the ID3 algorithm. Modifications to the entropy method are implemented to address various challenges in marketing campaign evaluation.

To achieve this objective, the following steps are undertaken:

– to modify the ID3 algorithm to enhance accuracy, prediction, and understanding of consumer response patterns in marketing campaigns, enabling more precise decision-making;

– to identify challenges related to consumer response imbalance to support more effective evaluation and adjustment of marketing campaign strategies;

– to address various technical challenges in decision tree implementation, such as overfitting, prediction errors, and data imbalance, to improve the reliability and accuracy of the model in analyzing market responses;

– to optimize data partitioning in the ID3 algorithm so that the resulting model is adaptive and accurate in effectively meeting market needs and demands;

– to provide real-case illustrations from the retail sector to demonstrate that the modified algorithm can be applied in practical scenarios.

4. Materials and methods

4. 1. Object and hypothesis of the study

The object of this study is the strategy of online retail marketing campaigns, particularly in the context of utilizing a modified ID3 decision tree algorithm to improve predictive effectiveness regarding consumer responses. The focus lies on how structural elements of campaigns – such as message framing, source credibility, and audience segmentation – affect consumer engagement and conversion rates.

The main hypothesis of this study is that the modification of the entropy calculation method in the ID3 algorithm can produce a more accurate and adaptive predictive model for identifying the most responsive audience segments in online retail marketing campaigns.

This research is based on several assumptions, including that historical campaign data reflect actual consumer behavior and that each attribute in the dataset contributes meaningfully to the classification process. Additionally, simplifications are made by excluding external variables such as macroeconomic conditions and competitor influences, in order to maintain analytical focus on the technical performance of the developed model.

4. 2. Modification of Shannon entropy

The simplification of the entropy formula [21], using logarithmic identities:

$$Ent(D)=-\frac{p}{p+n}\log_2\left(\frac{p}{p+n}\right)-\frac{n}{p+n}\log_2\left(\frac{n}{p+n}\right), \tag{1}$$

$$\log_2\left(\frac{\ln(x)}{\ln(b)}\right). \tag{2}$$

With b=2,

$$\log_2\left(\frac{p}{p+n}\right)=\frac{\ln\left(\frac{p}{p+n}\right)}{\ln(2)},$$

$$\log_2\left(\frac{n}{p+n}\right)=\frac{\ln\left(\frac{n}{p+n}\right)}{\ln(2)}.$$

Substitute into the entropy formula:

$$Ent(D)=-\frac{p}{p+n}\cdot\frac{\ln\left(\frac{p}{p+n}\right)}{\ln(2)}-\frac{n}{p+n}\cdot\frac{\ln\left(\frac{n}{p+n}\right)}{\ln(2)},$$

$$Ent(D)=\frac{1}{\ln(2)}\left[-\frac{p}{p+n}\ln\left(\frac{p}{p+n}\right)-\frac{n}{p+n}\ln\left(\frac{n}{p+n}\right)\right].$$

Combine terms within the logarithm:

$$-\frac{p}{p+n}\ln\left(\frac{p}{p+n}\right)-\frac{n}{p+n}\ln\left(\frac{n}{p+n}\right).$$

Rewrite the entropy terms using the natural logarithm (ln):

$$-\frac{p\ln(p)+n\ln(n)}{p+n}+\left(\frac{p\ln(p+n)+n\ln(p+n)}{p+n}\right).$$

Simplify further:

$$-\frac{p\ln(p)+n\ln(n)}{p+n}+\ln(p+n),$$

$$Ent(D)=\frac{1}{\ln(2)}\left[\ln(p+n)-\frac{p\ln(p)+n\ln(n)}{p+n}\right].$$

The above formula represents a modified entropy in information theory, utilized in the ID3 algorithm to measure uncertainty within a dataset. Entropy is calculated based on the probability of category occurrences in the data, represented by p_i. A high entropy value indicates a uniform distribution of categories, signifying a high degree of uncertainty in decision-making. Conversely, a low entropy value suggests that the data is concentrated in a single category, indicating a higher level of certainty. In decision modelling, entropy is used to identify the most informative attributes for decision tree construction. The attribute that results in the greatest entropy reduction after splitting is selected as the primary separator, enabling the model to classify data efficiently and accurately.

4. 3. Research method and algorithmic differences

To test this hypothesis, let's use two variants of the ID3 decision tree algorithm: the classical Shannon-based ID3 and a modified ID3 with a refined entropy function. Our focus was on how each algorithm partitions attributes (such as education, income, recency, etc.) to classify consumer responses (respond vs. not respond). The key differences are as follows:

1. Entropy calculation:

– classical ID3, uses the standard Shannon Entropy formula [13]:

$$Ent(D)=-\sum_{j=1}^{n}p_j\log_2 p_j, \tag{3}$$

where S represents the dataset under consideration, n denotes the total number of classes in the data, and p_j is the proportion or probability of elements belonging to the j-th class. The term $\log_2$ indicates the use of a base-2 logarithm, which reflects the concept of information in bits;

– modified ID3, employs a simplified or adjusted entropy term that reduces computational overhead and better handles skewed distributions by focusing on the natural logarithm of probabilities. This refinement seeks to improve classification when class imbalance is high.

2. Information gain and splitting criteria:

– classical ID3, selects attributes purely based on the largest reduction in Shannon entropy, this can overemphasize attributes with many distinct values;

– modified ID3, uses updated entropy in the information-gain calculation, aiming to mitigate multi-value bias and to balance the depth of the resulting tree against accuracy.

3. Handling of class imbalance:

– classical ID3, may bias splits toward the dominant (majority) class when the dataset is imbalanced;

– modified ID3, by adopting a more nuanced entropy formula, the algorithm can better identify minority-class patterns, improving recall for low-frequency outcomes.

Let's test both algorithms under varying data split ratios (e.g., 70:30 vs. 80:20) and tree depths (max_depth from 1 to 20). Our findings highlight that the modified ID3 yields higher accuracy and more stable performance than the classical version, particularly around a maximum tree depth of 6.

5. Results: factors influencing marketing campaigns

5. 1. Modified ID3

The simplified entropy formula is applied to the ID3 algorithm. The modified entropy formula [21], calculate dataset entropy, for dataset D with p and n:

$$Ent(D)=\frac{1}{\ln(2)}\left[\ln(p+n)-\frac{p\ln(p)+n\ln(n)}{p+n}\right]. \tag{4}$$

The variables in the formula are defined as follows: p represents the number of positive instances, n denotes the number of negative instances, and the sum of both, $p+n$, corresponds to the total number of instances in the dataset.

Calculate conditional entropy, attribute A with u distinct values, the dataset D is divided into subsets D_1, D_2, .., D_n:

$$Ent_A(D)=\sum_{j=1}^{u}\frac{|D_j|}{|D|}\left[\begin{array}{l}\ln(|D_j|)-\frac{p_j}{|D_j|}\ln(p_j)-\\-\frac{n_j}{|D_j|}\ln(n_j)\end{array}\right]. \quad (5)$$

The variable $|D_j|$ represents the number of elements in the subset D_j.

Additionally, p_j and n_j denote the number of positive and negative instances in the subset D_j, respectively. Calculate Information Gain, the Information Gain, Gain (A) is:

$$Gain(A)=Ent(D)-Ent_A(D). \quad (6)$$

If the $Gain(A)$ value is high, the attribute can significantly reduce uncertainty, making it ideal for use as a node in the decision tree. Conversely, if $Gain(A)$ is low, the attribute is less effective in distinguishing classes, meaning it does not contribute significantly to improving the model's accuracy. The information obtained from an attribute in a decision tree depends on the extent to which the attribute can reduce entropy in the dataset. The higher the information gain, the greater the attribute's influence in improving the quality of data partitioning, thereby supporting the development of a more robust model.

5. 2. Results of consumer response to marketing campaigns

The marketing campaign dataset contains data on 2,240 consumers, with 1,906 not responding to the campaigns and 334 providing responses. There is a significant difference between the number of responding and non-responding consumers, indicating an imbalance in the effectiveness of the marketing campaigns targeted at the audience. Examining data distribution is essential to understand patterns and factors that influence consumer response levels. The imbalance in proportions poses challenges for modeling data, such as the risk of bias toward the majority class, which can affect the accuracy and generalizability of predictive models.

The imbalance between responding and non-responding consumers can impact marketing strategies. Since the majority of consumers do not respond to the campaigns, companies need to reevaluate campaign elements using a data-driven approach. Consumer segmentation, personalized campaigns, communication channels, or timing adjustments could improve campaign effectiveness and appeal. These adjustments can help increase response rates by targeting consumer groups that are more specific and relevant. Additionally, analysis is needed to identify demographic or behavioral characteristics that distinguish consumers who respond from those who do not.

Fig. 1 illustrates that most consumers did not accept the marketing campaign offers. The largest number of rejections occurred across all campaign combinations from campaign_1 to campaign_5, with 1,631 consumers showing a response value of 0. The highest positive response was observed when all five campaigns were rejected, totaling 146 consumers. An interesting pattern emerged: positive responses were more frequent when one or two campaigns were accepted. In contrast, when multiple campaigns were accepted, such as campaign_5, the number of positive responses was only 21. Additionally, combinations of more than one campaign accepted-such as campaign_1 and campaign_4-generated positive responses from 10, 11, 12, and 14 consumers. These patterns indicate that campaign effectiveness is higher when not overwhelmed with too many simultaneous offers but focused on specific campaigns.

Positive responses tended to decline as the number of accepted campaigns increased. For example, only seven consumers responded positively when accepting campaigns 1, 2, and 4, indicating a decrease in consumer engagement. Overloading consumers with offers may lead to fatigue or confusion in making choices.

Fig. 1 highlights a significant difference in negative responses, with consumers rejecting all campaigns dominating the total responses. This shows that the majority of consumers are not interested in the offers provided, signaling the need for more focused and personalized strategies in delivering campaigns [22]. These strategies could increase consumer engagement and improve conversion rates.

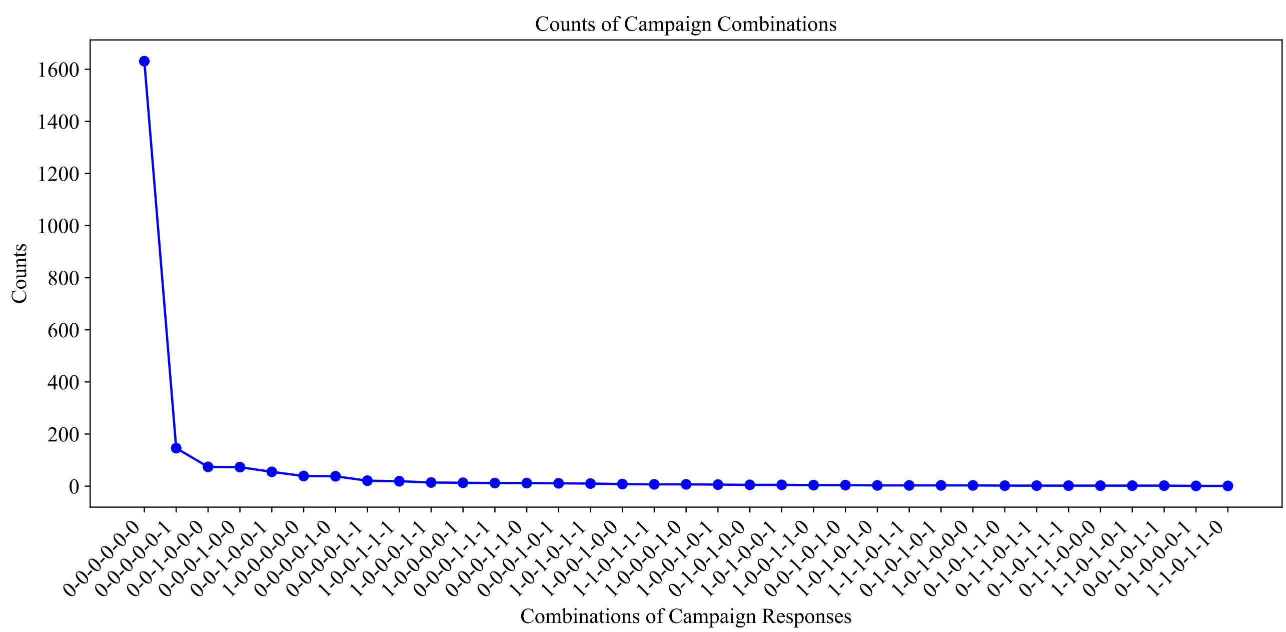

Fig. 1. Distribution of consumer responses to marketing campaigns

Fig. 1 highlights a significant difference in negative responses, with consumers rejecting all campaigns dominating the total responses. This shows that the majority of consumers are not interested in the offers provided, signaling the need for more focused and personalized strategies in delivering campaigns [15]. These strategies could increase consumer engagement and improve conversion rates.

In the context of marketing campaigns, the results underscore the importance of understanding consumer preferences and behavior. Positive responses are higher when only one or two campaigns are offered. Companies can tailor their approach to deliver campaigns that are more relevant and aligned with consumer needs. Focusing on quality rather than quantity is an effective strategy for enhancing consumer engagement and satisfaction [15].

Fig. 2 illustrates that the highest response rate is observed among consumers with an income range of 100,000 to <300,000, with a response rate of 0.3333. Consumers with an income between 50,000 and <100,000 exhibit a response rate of 0.117, while those in the 5,000 to <10,000 range have a response rate of 0.130. Consumers with an income between 1,730 and <5,000, as well as those earning more than 500,000, do not respond, indicating low engagement or a lack of interest in marketing campaigns.

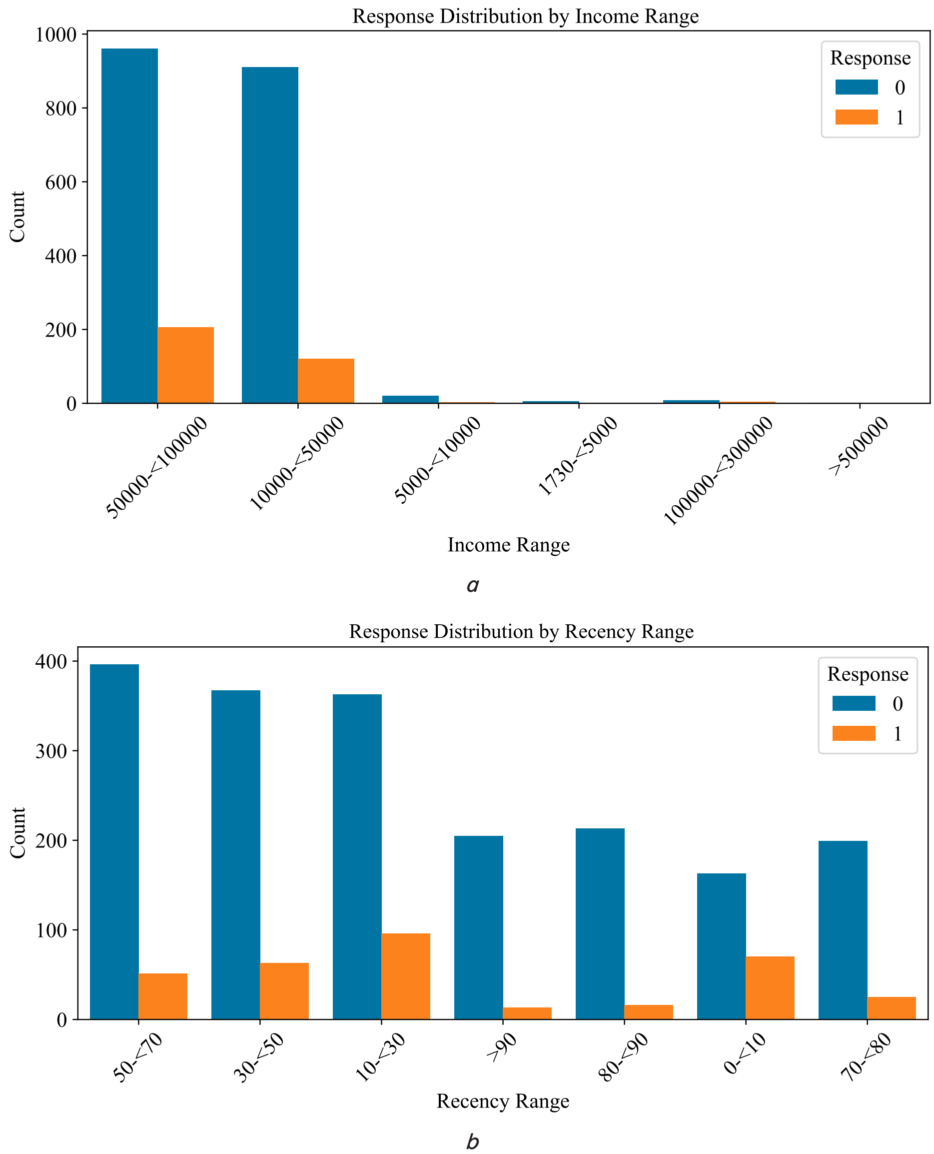

Fig. 2. Distribution of campaign responses: *a* – based on income range; *b* – based on recency range

Consumer interaction significantly influences response rates in marketing campaigns. Consumers who recently interacted within 0 to <10 days exhibit the highest response rate at 0.300, followed by those in the 10 to <30-day range with a response rate of 0.209 and those in the 30 to <50-day range with a response rate of 0.147. Consumers with longer interaction periods, specifically those in the 80 to <90-day and >90-day categories, display lower response rates of 0.070 and 0.060, respectively. This indicates that the longer the time since the last interaction, the lower the likelihood of consumers responding to marketing campaigns.

The second chart illustrates response distribution by recency, where consumers who recently engaged with the campaigns or the company show higher response rates, both positive and negative. This demonstrates that the last interaction with the consumer can influence their attitude toward the campaigns. Newly engaged consumers may be open to new offers or find the campaigns more relevant to their current situation [23]. Meanwhile, consumers who have not interacted with the company for a long time tend to reject campaigns, likely due to a lack of engagement or interest in the offered promotions.

When combining insights from both charts, a correlation emerges between income range and recency in influencing consumer responses to marketing campaigns. This indicates that effective marketing strategies should align with the characteristics of consumer segmentation [24]. For high-income, recently engaged consumers, companies could offer premium products or services with a more personalized approach. For low-income, disengaged consumers, companies should tailor their offers in a way that rekindles their interest and engagement, ensuring relevance and appeal to their specific needs.

In Fig. 3, individuals with higher educational backgrounds tend to be more responsive to marketing campaigns compared to those with lower education levels. Consumers with advanced academic degrees demonstrate active engagement with marketing messages, which can be attributed to their higher level of understanding and awareness of the value of the promoted products or services. Conversely, individuals with lower education levels are less likely to respond, possibly due to limited access to information or a lack of campaign relevance to their needs. This difference indicates that effective marketing strategies should be tailored to the characteristics of the audience based on their education level, ensuring that messages are easily understood and well-received across various consumer segments.

Fig. 3 presents the distribution of responses based on education level, revealing interesting results. Consumers with a PhD recorded the highest response rate at 20.7 %, indicating that higher education levels correspond to greater responsiveness to marketing campaigns [22]. This could be attributed to their higher level of information and awareness regarding the importance of the promoted products or services. It may also reflect their maturity in making purchase decisions, enabling quicker responses to offers.

Consumers with a master's degree showed a response rate of 15.4 %, highlighting a substantial level of engagement in marketing campaigns among individuals with higher education backgrounds. Lower education groups, such as Graduation at 13.4 %, 2nd Cycle at 10.8 %, and Basic at 3.7 %, exhibited lower response rates. These findings suggest variations in knowledge or needs related to the promoted products. Companies might need to tailor their campaign messages and approaches to align with the education levels and understanding of different audience segments.

The data suggests further analysis is needed to identify factors in marketing campaigns that influence consumer responses [16]. The type of product or service being promoted is closely related to the education level of consumers. Factors such as education, income, recency, and prior experiences with the brand also play a role in shaping response outcomes.

An evaluation of key elements in online retail marketing highlights that market framing the way messages are structured and presented plays a crucial role in campaign effectiveness. The data indicates that the majority of consumers did not respond to the campaigns, suggesting that the framing strategies employed have not fully captured audience interest. Overly frequent or misaligned campaigns may lead to information fatigue, thereby reducing their effectiveness. A more personalized approach in crafting marketing messages is needed to improve consumer engagement and campaign appeal.

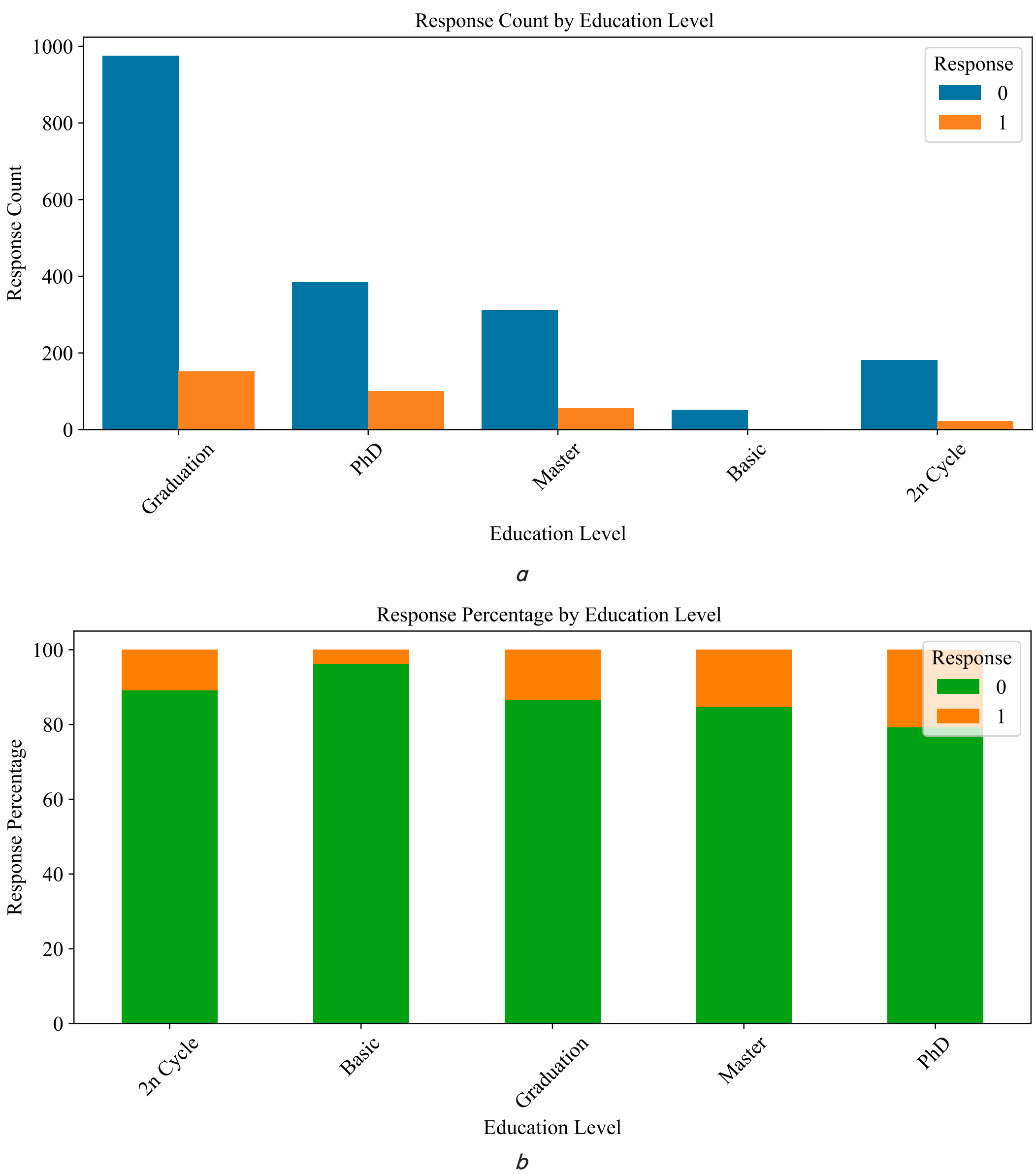

Fig. 3. Educational responses: *a* — based on education level; *b* — percentage of education levels

The credibility of the information source in marketing campaigns significantly influences consumer decision-making. The data reveals that positive responses are more likely when campaigns are presented selectively, rather than overwhelming consumers with too many promotions at once. Consumers tend to place greater trust in well-structured and less aggressive campaigns. Moreover, educational level continues to impact response rates PhD holders, for instance, recorded the highest response at 20.7 % (Fig. 3, *a*), suggesting that highly educated consumers tend to be more selective in evaluating the credibility of information sources before making decisions.

Audience segmentation is a critical factor in the effectiveness of online retail marketing campaigns. The data analysis shows that the highest response rate came from consumers with an income range between 100,000 and <300,000, with a response rate of 0.333 (Fig. 2, *a*). Consumers with either lower or higher incomes displayed lower response rates. Furthermore, recent interaction with the company influenced consumer engagement, with the highest responses recorded among consumers who had interacted within the past 0 to <10 days (Fig. 2, *b*). This highlights the need for marketing strategies that consider segmentation based on both income and engagement levels.

The combination of appropriate audience segmentation and effective market framing strategies can enhance marketing campaign effectiveness. The data shows that consumers who received only one or two campaign messages demonstrated higher response rates compared to those exposed to multiple simultaneous campaigns. A targeted approach that aligns with audience needs can significantly boost consumer engagement. Conversely, excessive exposure to campaigns may lead to confusion or fatigue, ultimately reducing the overall impact of the strategy.

In summary, the evaluation of key elements in online retail marketing demonstrates that market framing, information source credibility, and audience segmentation must be optimized in an integrated manner. The available data suggests that marketing strategies should be more personalized, based on clear segmentation, and supported by credible information to enhance consumer response rates. By adjusting strategies in line with these factors, companies can develop more effective marketing campaigns, strengthen consumer engagement, and achieve higher conversion rates.

5. 3. Adapting ID3 model concepts

The implementation of marketing campaigns using the ID3 model can be effective in analyzing and understanding consumer behavior patterns. By utilizing variables such as education level, marital status, children at home, and teenagers at home, the model identifies the most influential attributes affecting purchasing decisions. ID3 explores how education level and marital status impact the likelihood of consumers accepting offers and making purchases. Variables like income range and recency provide insights into customer segments, highlighting tendencies to purchase based on income level and recent interactions with the company. New consumers who interact through promotions or customer service are more likely to make a purchase compared to those who have not engaged with online retail for a long time.

ID3 leverages consumer behavior data, such as the number of purchases made online, through catalogs, and in physical stores. The decision tree model identifies the most effective distribution channels for various market segments by analyzing purchase frequency through these channels. This allows companies to optimize marketing strategies to align with consumer shopping preferences. Variables such as expenditure ranges for wine, fruit, and meat reveal consumer spending habits in specific product categories, providing valuable information for tailoring product offerings to be more relevant.

The application of the ID3 method, through adaptive attribute utilization, enables companies to sign marketing campaigns that are segmented and tailored to the characteristics of specific consumer groups. The decision tree reveals that consumers with teenagers at home and high income are particularly responsive to offers for wine and sweet foods. Companies can adjust their marketing campaigns to target these segments with more specific and appealing messages. By applying the ID3 method, a deeper understanding of the factors influencing consumer decisions can significantly enhance the effectiveness of marketing campaigns [15].

In Table 1, the ID3 algorithm using an 80:20 data split achieves an accuracy of 0.8415 with a maximum tree depth of 5 and a computation time of 0.6075 seconds. The precision for class 0 is 0.85, while for class 1, it is 0.43. The recall for class 0 is 0.98, and for class 1, it is 0.09. The F1-score for class 0 is 0.91, indicating good performance in predicting the majority class. However, the F1-score for class 1 is only 0.14, reflecting inadequate performance in predicting the minority class.

Table 1

The impact of data split ratios on the performance of the modified ID3 algorithm

Split data	Precision		Recall		F1-score		Accuracy	Time (second)
	0	1	0	1	0	1		
80:20	0.85	0.43	0.98	0.09	0.91	0.14	0.8415178571428571	0.6074917316436
70:30	0.87	0.45	0.98	0.09	0.92	0.16	0.8556547619047619	3.1733415126800

With a 70:30 data split and a tree depth of max_depth 5, accuracy improves to 0.8557, although the computation time increases to 3.1733 seconds. The precision for class 0 rises to 0.87, higher than that achieved with the 80:20 split, while the precision for class 1 is 0.45. The recall for class 0 remains at 0.98, and for class 1, it stays low at 0.09. The F1-score for class 0 increases to 0.92, demonstrating consistent performance in predicting the majority class. Meanwhile, the F1-score for class 1 improves slightly to 0.16, highlighting the ongoing challenge of handling data imbalance in the minority class. The results of the metric calculations and their impact on the model's performance can be observed in Fig. 4.

Fig. 4 compares the performance metrics of the Modified ID3 algorithm using two different data splits. The first chart illustrates model accuracy, where the 70:30 split achieves slightly higher accuracy (0.8557) compared to the 80:20 split (0.8415). The 80:20 split produces a high precision for class 0 (0.85), while precision for class 1 remains low for both splits. The recall metric shows the model performs well in predicting class 0, achieving a value of 0.98 for both splits, but recall for class 1 remains consistently low at 0.09. The F1-score follows a similar trend, with class 0 achieving high values that indicate good performance, while class 1 shows low F1-scores, reflecting the difficulty of handling the minority class. The final chart highlights differences in computation time. The 80:20 split achieves faster computation time at 0.6075 seconds, whereas the 70:30 split requires significantly longer computation time at 3.1733 seconds.

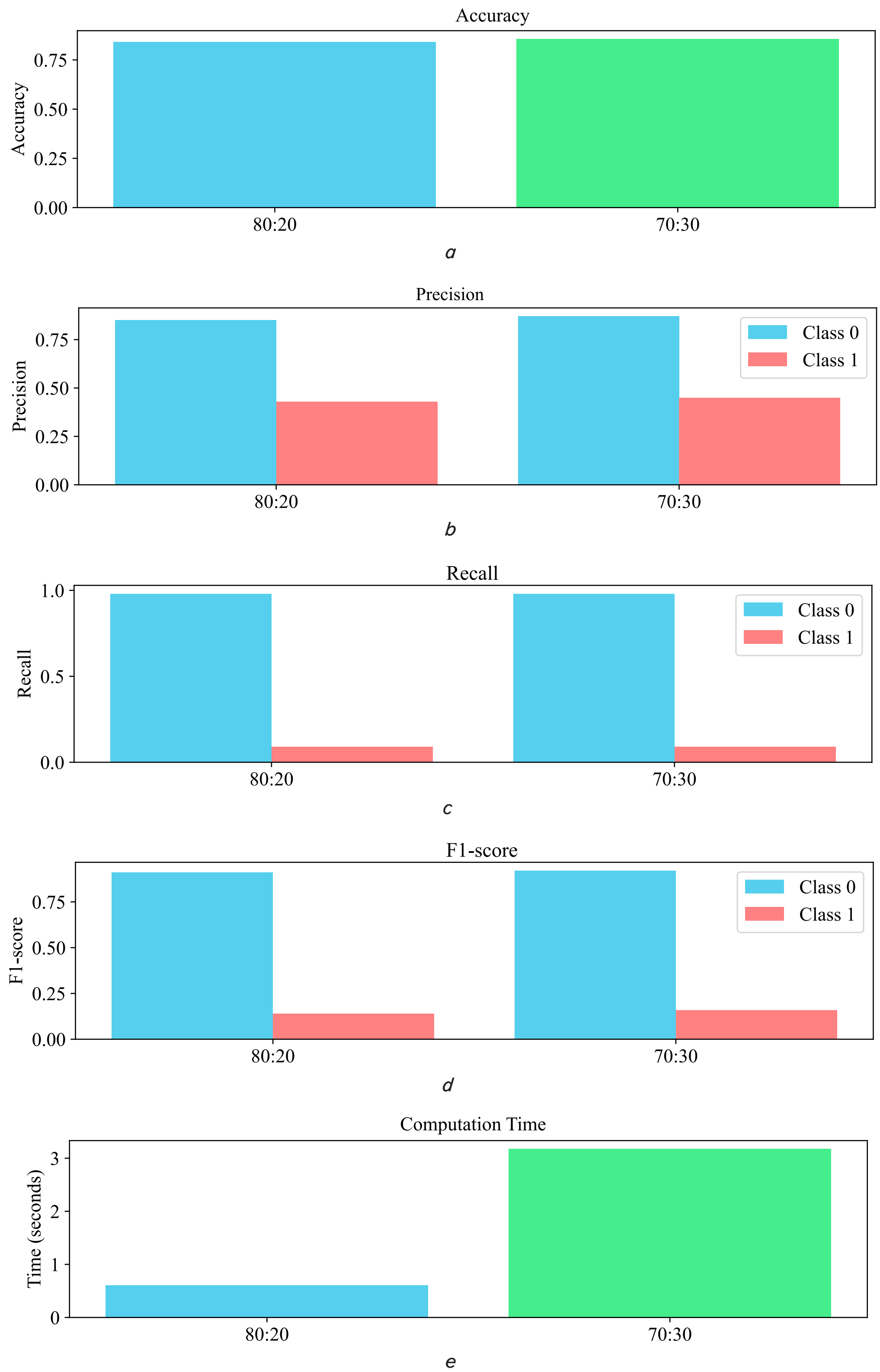

Fig. 4. Performance of the modified ID3 model based on data split ratio: *a* — accuracy; *b* — precision; *c* — recall; *d* — F1-score; *e* — computation time

5. 4. Adapting model concepts and tree depth

In decision tree modelling using ID3, the parameter max_depth is crucial for controlling model complexity and avoiding overfitting [25]. Max_depth refers to the maximum depth a decision tree can reach during the training process. Tree depth determines the number of splits that can occur at each level of the data [26]. Deeper trees allow more splits, enabling the model to capture more patterns or relationships within the data. However, if the depth is too large, the model may become overly specific to the training data, reducing its ability to generalize to new data.

The selection of the best attribute for each split is based on information gain, which measures the reduction in entropy after the split. Introducing a limit on max_depth restricts the number of splits. thereby reducing model complexity and mitigating overfitting. Conversely, if max_depth is too small, the model may lack the complexity needed to capture patterns in the data, leading to underfitting [27]. Selecting an appropriate value for max_depth is essential to achieve a balance between bias and variance. This ensures the creation of an accurate model that generalizes well to unseen data.

Table 2 demonstrates that max_depth has a significant impact on model accuracy and execution time. For max_depth values between 1 and 3, accuracy remains stable at 0.8586, while execution time increases from 0.0778 seconds to 0.2714 seconds. Between max_depth 4 and 6, there is slight fluctuation in accuracy, with max_depth 6 achieving the highest accuracy of 0.8616 and an execution time of 0.7542 seconds. Beyond max_depth 6, accuracy gradually decreases, and execution time increases significantly. At max_depth 10, accuracy drops to 0.8304, while execution time rises to 2.0241 seconds. This trend continues up to max_depth 20, where accuracy reaches 0.7887 and execution time peaks at 3.7831 seconds. Overall, increasing max_depth allows the model to capture more data complexity. However, it also leads to longer execution times and potential overfitting, as evidenced by the decline in accuracy beyond a certain threshold.

Fig. 5 illustrates the testing results for various decision tree depths, ranging from max_depth 1 to 20. At lower tree depths, specifically between max_depth 1 and 6, accuracy remains high, with the highest accuracy of 0.8616 achieved at max_depth 6. Computation time at lower depths is relatively short, starting at 0.0778 seconds for max_depth 1 and gradually increasing as tree depth grows.

At max_depth 7, a significant drop in accuracy is observed, accompanied by a longer computation time. Accuracy continues to decline gradually after max_depth 6, with max_depth 9 producing the lowest accuracy of 0.8304 and a computation time of 2.0241 seconds. As tree depth increases further, accuracy declines further, and computation time becomes significantly longer, reaching 4.5280 seconds at max_depth 19.

Table 2

The impact of max depth on model accuracy and execution time

Max_depth	Accuracy	Time (seconds)
1	0.8586	0.0778
2	0.8586	0.1598
3	0.8586	0.2714
4	0.8557	0.4208
5	0.8557	0.5436
6	0.8616	0.7542
7	0.8586	1.0438
8	0.8497	1.3326
9	0.8348	1.6570
10	0.8304	2.0241
11	0.8155	2.8989
12	0.8006	3.5023
13	0.7917	2.8825
14	0.7932	3.1021
15	0.7887	3.3501
16	0.7887	4.3187
17	0.7887	3.4238
18	0.7887	3.3925
19	0.7887	4.5280
20	0.7887	3.7831

5. 5. Illustrative real-case examples

To demonstrate practical applicability, let's briefly present two real-case scenarios illustrating how the modified ID3 decision tree can inform marketing strategies:

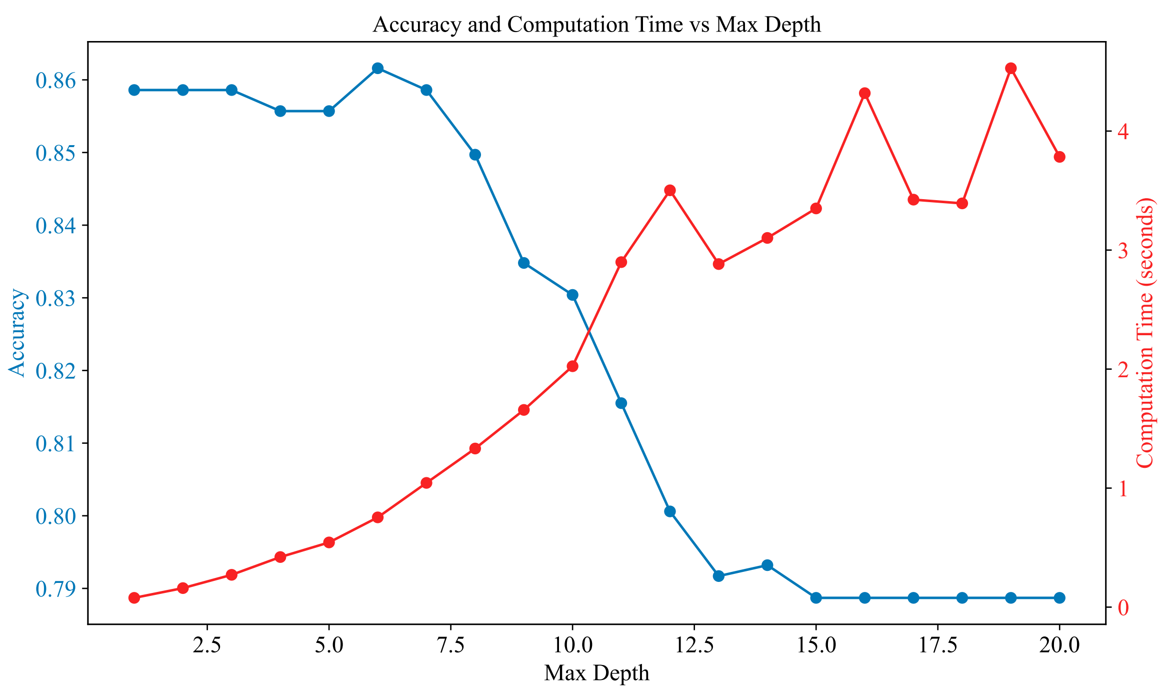

Fig. 5. Accuracy and computation time vs max depth

1. Retail apparel campaign, in a pilot project with a regional apparel retailer, a marketing team wanted to optimize email promotions for new arrivals and seasonal discounts:

– consumer segment, the retailer noticed that among higher-income consumers (incomes>100,000 USD), those who had interacted with the website within the last 10 days were significantly more inclined to open promotional emails and use discount coupons;

– ID3 insights, the modified ID3 tree identified "recency of interaction" and "education level" as the top two splitting attributes, leading the retailer to send targeted offers on premium lines to this segment;

– outcome: this focused campaign (only one or two offers per email) increased email open rates by 15 % and boosted conversions by 10 % compared to blanket promotions previously offered.

2. Online grocery promotions, an online grocery service tested multiple campaign variations (e.g., discounts on fresh produce vs. packaged goods):

– consumer segment, households with teenagers and a moderate income (50k–100k USD) often responded best when offered discounts on fruit and meat-likely due to family meal-planning habits;

– ID3 insights, the modified ID3 tree highlighted that presenting too many product deals at once (e.g., fruit, dairy, and snacks simultaneously) reduced acceptance rates (choice overload). The tree indicated that one or two well-targeted deals boosted redemption rates;

– outcome, by limiting offers to two per email, the grocery service reported 20 % higher redemption for fruit and meat promotions and a notable drop in campaign unsubscribes.

Through these real-case illustrations, it is possible to show how the modified ID3 algorithm helps pinpoint the most influential attributes (income, recency, education, and so forth) and reveals that fewer, more targeted offers improve consumer response rates. Such insights align with our hypothesis that optimizing the campaign structure – rather than merely adding more offers – can significantly enhance marketing efficiency.

6. Discussion of the results of the study comparison of marketing campaign performance based on consumer responses

The study reveals that the effectiveness of marketing campaigns is influenced by the number of offers received by consumers. The evaluation results indicate that campaigns with one or two offers have a higher acceptance rate compared to campaigns that present multiple offers simultaneously. The phenomenon of choice overload [28] is clearly evident in the data, where an excessive number of choices hinders decision-making and reduces consumer responses. As shown in Fig. 1, a simple marketing strategy, such as a single campaign (campaign_5) or a combination of two campaigns (campaign_1 and campaign_4), has proven effective in capturing consumer attention compared to strategies that include three or more campaigns. The decline in responses to complex campaign combinations suggests that consumers tend to reject or ignore offers when confronted with too many options at once, reinforcing the notion that a more focused marketing approach can enhance campaign success.

The evaluation of decision tree depth shows that models with lower depth (max_depth between 1 and 6) maintain stable accuracy and efficient computation time. As presented in Table 2, models with max_depth between 1 and 6 demonstrate stable accuracy above 80 %. However, increasing the depth beyond max_depth 6 leads to a decrease in accuracy below 75 % and an increase in processing time of more than 50 %. Fig. 5 confirms that the deeper the decision tree, the more complex the model becomes, resulting in overfitting and longer computation times. These results align with previous research on the impact of max_depth on the Random Forest algorithm [29]. A model with max_depth 1 achieved 97 % accuracy, but as the tree depth increased to 20, accuracy dropped to 90 %. This finding suggests that while greater depth allows the model to capture more patterns in the data, excessive increases can degrade overall model performance due to overfitting.

The proposed advantage lies in a response pattern-based approach that identifies campaign effectiveness based on the number of offers given, rather than merely considering the overall campaign success rate, as illustrated in Fig. 1. There is a clear relationship between the number of campaigns accepted and consumers' positive responses. Campaigns with two offers exhibit a higher positive response rate compared to those with three or more offers. This approach highlights the importance of balancing the number of offers presented to consumers to avoid the effects of choice overload. The evaluation of the modified ID3 algorithm in determining the optimal decision tree depth provides valuable insights into model optimization, ensuring that accuracy is maintained without excessively increasing computational complexity.

The obtained solution demonstrates that campaign effectiveness does not solely depend on the quality of offers but also on the number of campaigns presented to consumers. Addressing the issue of low campaign acceptance, the study confirms that a more focused marketing strategy can enhance consumer engagement. From a modeling perspective, this reinforces the importance of determining the optimal decision tree depth to maintain model efficiency while avoiding overfitting. As shown in Fig. 5, the optimal depth yields accurate classification with reduced computation time compared to deeper models. Compared to the Random Forest algorithm [29], using a lower tree depth proves to be more optimal in maintaining accuracy than excessive depth, which leads to performance degradation.

The limitations of this study lie in the scope of the data used, which may not fully represent all market segments. The decision tree model also has limitations in handling complex data with high variable interactions. The application of this model must consider the specific characteristics of the dataset used. One of the weaknesses of this study is the lack of exploration into psychological factors influencing consumer decisions to accept or reject campaigns. While the findings indicate a relationship between the number of campaigns accepted and positive responses, the study does not delve deeper into the specific reasons behind consumer decisions. Additionally, the modified ID3 model remains limited to decision tree-based evaluation without comparisons to other methods that might be more effective in handling data complexity.

Future research should focus on addressing class imbalance issues in consumer responses to marketing campaigns. Algorithms such as Synthetic Minority Over-sampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), or cost-sensitive learning can be implemented to improve the representation of minority classes in the dataset. Testing with a broader and more diverse dataset is also necessary to ensure the applicability of these findings across various industrial contexts.

The study reveals that the effectiveness of marketing campaigns is influenced by the number of offers received by consumers. The evaluation results indicate that campaigns with one or two offers have a higher acceptance rate compared to those presenting multiple offers simultaneously. This suggests the presence of a choice overload effect, where an excessive number of options hinders decision-making and reduces consumers' positive responses. As shown in Fig. 1, a simpler marketing strategy, such as a single campaign (campaign_5) or a combination of two campaigns (campaign_1 and campaign_4), is more effective in attracting consumer attention compared to strategies involving three or more campaigns. The decline in response rates for complex campaign combinations suggests that consumers tend to reject or ignore offers when confronted with too many choices at once, reinforcing the notion that a more focused marketing approach enhances campaign success.

The evaluation of decision tree depth indicates that models with a lower depth (max_depth between 1 and 6) maintain stable accuracy and efficient computational time. As presented in Table 2, models with max_depth between 1 and 6 exhibit stable accuracy above 80 %. However, increasing the depth beyond max_depth 6 results in an accuracy decline below 75 % and an increase in processing time by more than 50 %. Fig. 5 confirms that as the decision tree grows deeper, the resulting model becomes more complex, leading to overfitting and prolonged computational time.

The advantage of the proposed solution lies in its response pattern-based approach, which identifies campaign effectiveness based on the number of offers provided rather than solely considering aggregate campaign success. As illustrated in Fig. 1, there is a clear relationship between the number of accepted campaigns and the rate of positive responses. Campaigns with two offers achieve a higher positive response rate than those with three or more offers. This approach highlights the importance of balancing the number of offers presented to consumers to avoid choice overload. The evaluation of the modified ID3 algorithm in determining the optimal decision tree depth provides valuable insights into model optimization, ensuring that accuracy is maintained without increasing computational complexity.

The obtained solution provides evidence that campaign effectiveness is not solely dependent on the quality of the offers but also on the number of campaigns presented to consumers. This addresses the issue of low campaign acceptance rates, demonstrating that a focused marketing strategy can enhance consumer engagement. From a modeling perspective, these findings reinforce the importance of determining the optimal decision tree depth to maintain model efficiency while avoiding overfitting. As depicted in Fig. 5, the decision distribution at the optimal depth produces accurate classification results, maintaining a lower number of nodes and reduced computational time compared to models with excessive depth.

The limitations of this study lie in the scope of the data used, which may not represent all market segments. Additionally, the ID3 decision tree model has constraints in handling complex data with high variable interactions. The applicability of these findings must consider the specific characteristics of the dataset used. One of the shortcomings of this study is the lack of exploration into psychological factors influencing consumer decisions to accept or reject campaigns. Although the findings indicate a relationship between the number of accepted campaigns and positive responses, the study does not delve deeper into the specific reasons behind consumer decisions.

The modified ID3 model remains limited to decision tree-based evaluation without comparisons to other methods that might be more effective in handling data complexity. Future research can focus on addressing class imbalance issues in consumer responses to marketing campaigns. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), or cost-sensitive learning can be implemented to improve the representation of minority classes in the dataset. Testing with a more extensive and diverse dataset can help ensure that these findings are applicable across various industry contexts.

The results indicate that campaign acceptance is higher when consumers receive one or two offers compared to multiple offers simultaneously. The phenomenon of choice overload, where an excessive number of choices leads to confusion and decision fatigue, is evident. The data show that simple campaign combinations, such as campaign_5 or the combination of campaign_1 and campaign_4, have higher acceptance rates compared to combinations involving three or more campaigns, which tend to experience a decline in positive responses.

The evaluation of decision tree depth reveals that models with a lower tree depth (max_depth between 1 and 6) maintain stable accuracy and efficient computational time. However, increasing the depth beyond max_depth 6 leads to a decrease in accuracy and an increase in processing time.

The advantage of the proposed solution lies in its response pattern-based approach, which identifies campaign effectiveness based on the number of offers provided, rather than solely considering overall campaign success rates in aggregate. This approach highlights the relationship between the number of accepted campaigns and positive responses. Additionally, the evaluation of the ID3 algorithm in determining the optimal decision tree depth provides valuable insights into model optimization, ensuring accuracy is maintained without increasing computational complexity.

The obtained solution provides evidence that campaign effectiveness does not solely depend on the quality of the offers but also on the number of campaigns presented to consumers. This addresses the issue of low campaign acceptance rates, demonstrating that a more focused marketing strategy can enhance consumer engagement. From a modelling perspective, it reinforces the importance of determining the optimal decision tree depth to keep the model efficient while avoiding overfitting.

The study's limitations lie in the scope of the data used, which may not represent all market segments. Additionally, the ID3 decision tree model has constraints in handling complex data with high variable interactions. The applicability of the findings in other contexts must consider the specific characteristics of the dataset used.

The shortcomings of this study include the lack of exploration of psychological factors influencing consumer decisions to accept or reject campaigns. While the results indicate a relationship between the number of accepted campaigns and positive responses, the study does not delve deeper into the specific reasons behind consumer decisions. Furthermore, the modified ID3 model used is still limited to decision tree-based evaluation without comparison to other methods that may be more effective in handling data complexity.

Future research can focus on addressing class imbalance issues in consumer responses to marketing campaigns. Techniques such as Synthetic Minority Over-sampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), or cost-sensitive learning can be implemented to improve the representation of minority classes in the dataset. An ensemble learning-based approach, combining Random Forest and Gradient Boosting with data balancing techniques, can enhance accuracy in highly imbalanced class distributions. Additionally, testing with a more extensive and diverse dataset can help ensure that these findings are applicable across various industry contexts.

7. Conclusions

1. The proposed model aims to improve accuracy and predictive capability in understanding consumer response patterns to online marketing campaigns. It modifies the ID3 algorithm by applying a simplified entropy formula better suited for imbalanced data. This model is expected to support more precise and effective decision-making in marketing strategy planning.

2. The success of marketing campaigns is influenced by education, income, and consumer interaction. The highest response rate of 20.7 % came from consumers with a PhD, those with an income between 100,000 and <300,000 showed a 33 % response rate, and those who interacted within 0 to <10 days had a 30 % response rate. Positive responses declined when too many campaigns were accepted, indicating that selective and relevant strategies are more effective in increasing response rates and conversions.

3. ID3 demonstrates a high accuracy of 85.57 % in the 70:30 data split scenario; however, its performance on the minority class remains low, with an F1-score of 0.16. This indicates the need for addressing data imbalance to improve predictive performance on consumers who provide a positive response.

4. Increasing the max_depth enhances the model's ability to capture data complexity; however, beyond max_depth 6, accuracy decreases while execution time increases. The optimal value is observed at max_depth 6, with an accuracy of 0.8616 and an execution time of 0.7542 seconds, indicating the need to balance model complexity and computational efficiency.

5. The real-case illustrations demonstrate that the modified ID3 algorithm is capable of identifying key attributes and show that fewer, more targeted offers improve consumer response, supporting the optimization of campaign structure.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Manuscript has data included as electronic supplementary material.

Use of artificial intelligence

The authors have used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

References

1. Byrne, S., Hart, P. S. (2009). The Boomerang Effect A Synthesis of Findings and a Preliminary Theoretical Framework. Annals of the International Communication Association, 33 (1), 3–37. https://doi.org/10.1080/23808985.2009.11679083
2. Brun, A., Castelli, C. (2008). Supply chain strategy in the fashion industry: Developing a portfolio model depending on product, retail channel and brand. International Journal of Production Economics, 116 (2), 169–181. https://doi.org/10.1016/j.ijpe.2008.09.011
3. Kaiser, C., Ahuvia, A., Rauschnabel, P. A., Wimble, M. (2020). Social media monitoring: What can marketers learn from Facebook brand photos? Journal of Business Research, 117, 707–717. https://doi.org/10.1016/j.jbusres.2019.09.017
4. Wettstein, D., Suggs, L. S. (2016). Is it social marketing? The benchmarks meet the social marketing indicator. Journal of Social Marketing, 6 (1), 2–17. https://doi.org/10.1108/jsocm-05-2014-0034
5. Erkan, I., Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. Computers in Human Behavior, 61, 47–55. https://doi.org/10.1016/j.chb.2016.03.003
6. Wang, X., Yu, C., Wei, Y. (2012). Social Media Peer Communication and Impacts on Purchase Intentions: A Consumer Socialization Framework. Journal of Interactive Marketing, 26 (4), 198–208. https://doi.org/10.1016/j.intmar.2011.11.004
7. Gunawan, D. D., Huarng, K.-H. (2015). Viral effects of social network and media on consumers' purchase intention. Journal of Business Research, 68 (11), 2237–2241. https://doi.org/10.1016/j.jbusres.2015.06.004
8. Wandosell, G., Parra-Meroño, M. C., Alcayde, A., Baños, R. (2021). Green Packaging from Consumer and Business Perspectives. Sustainability, 13 (3), 1356. https://doi.org/10.3390/su13031356
9. Pace, K., Silk, K., Nazione, S., Fournier, L., Collins-Eaglin, J. (2016). Promoting Mental Health Help-Seeking Behavior Among First-Year College Students. Health Communication, 33 (2), 102–110. https://doi.org/10.1080/10410236.2016.1250065
10. Melović, B., Dabić, M., Vukčević, M., Ćirović, D., Backović, T. (2021). Strategic business decision making: the use and relevance of marketing metrics and knowledge management. Journal of Knowledge Management, 25 (11), 175–202. https://doi.org/10.1108/jkm-10-2020-0764

11. López, V., Fernández, A., García, S., Palade, V., Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Information Sciences, 250, 113–141. https://doi.org/10.1016/j.ins.2013.07.007
12. Horton, S., Towell, M., Haegeli, P. (2020). Examining the operational use of avalanche problems with decision trees and model-generated weather and snowpack variables. Natural Hazards and Earth System Sciences, 20 (12), 3551–3576. https://doi.org/10.5194/nhess-20-3551-2020
13. Wang, Y., Li, Y., Song, Y., Rong, X., Zhang, S. (2017). Improvement of ID3 Algorithm Based on Simplified Information Entropy and Coordination Degree. Algorithms, 10 (4), 124. https://doi.org/10.3390/a10040124
14. Chen, Z., Ji, J. (2023). Analysis and Identification of the Composition of Ancient Glass Objects Based on Logistic Regression Analysis. Highlights in Science, Engineering and Technology, 41, 265–270. https://doi.org/10.54097/hset.v41i.6830
15. Middleton, L., Dowdle, D., Villa, L., Gray, J., Cumming, J. (2019). Saving 20 000 Days and Beyond: a realist evaluation of two quality improvement campaigns to manage hospital demand in a New Zealand District Health Board. BMJ Open Quality, 8 (4), e000374. https://doi.org/10.1136/bmjoq-2018-000374
16. Giovanis, A. N., Athanasopoulou, P. (2018). Consumer-brand relationships and brand loyalty in technology-mediated services. Journal of Retailing and Consumer Services, 40, 287–294. https://doi.org/10.1016/j.jretconser.2017.03.003
17. Kranzler, E. C., Gibson, L. A., Hornik, R. C. (2017). Recall of "The Real Cost" Anti-Smoking Campaign Is Specifically Associated With Endorsement of Campaign-Targeted Beliefs. Journal of Health Communication, 22 (10), 818–828. https://doi.org/10.1080/10810730.2017.1364311
18. Huang, K., Wang, T. (2024). Optimized Application of the Decision Tree ID3 Algorithm Based on Big Data in Sports Performance Management. International Journal of E-Collaboration, 20 (1), 1–20. https://doi.org/10.4018/ijec.350022
19. Deniz, N. (2020). Cognitive biases in MCDM methods: an embedded filter proposal through sustainable supplier selection problem. Journal of Enterprise Information Management, 33 (5), 947–963. https://doi.org/10.1108/jeim-09-2019-0285
20. Sun, H., Hu, X. (2017). Attribute selection for decision tree learning with class constraint. Chemometrics and Intelligent Laboratory Systems, 163, 16–23. https://doi.org/10.1016/j.chemolab.2017.02.004
21. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1 (1), 81–106. https://doi.org/10.1007/bf00116251
22. Narayanan, S., Das, J. R. (2021). Can the marketing innovation of purpose branding make brands meaningful and relevant? International Journal of Innovation Science, 14 (3/4), 519–536. https://doi.org/10.1108/ijis-11-2020-0272
23. Kaplan, A. M., Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons, 53 (1), 59–68. https://doi.org/10.1016/j.bushor.2009.09.003
24. Souiden, N. (2002). Segmenting the Arab markets on the basis of marketing stimuli. International Marketing Review, 19 (6), 611–636. https://doi.org/10.1108/02651330210451944
25. Chu, W., Qu, X. (2024). Severitys prediction of car accidents in PA and model comparison. Applied and Computational Engineering, 52 (1), 215–226. https://doi.org/10.54254/2755-2721/52/20241582
26. Lee, S., Lee, S., Park, Y. (2007). A prediction model for success of services in e-commerce using decision tree: E-customer's attitude towards online service. Expert Systems with Applications, 33 (3), 572–581. https://doi.org/10.1016/j.eswa.2006.06.005
27. Hancock, J. T., Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. Journal of Big Data, 7 (1). https://doi.org/10.1186/s40537-020-00369-8
28. Papadopoulou, N., Raïes, K., Mir Bernal, P., Woodside, A. G. (2019). Gifts as conduits in choice overload environments. Psychology & Marketing, 36 (7), 716–729. https://doi.org/10.1002/mar.21207
29. Vins, A. D. S., Sam Emmanuel, W. R. (2021). Optimized Random Forest Algorithm with Parameter Tuning for Predicting Heart Disease. Advances in Computing and Data Sciences, 443–451. https://doi.org/10.1007/978-3-030-81462-5_40