Predicting Campaign ROI Using Decision Trees and Random Forests in Digital Marketing

B Herawan Hayadi[1,*], Ibrahiem M. M. El Emary[2]

[1]Primary School Teacher Education,Universitas Bina Bangsa, Serang, Indonesia

[2]King Abdulaziz University, Kingdom of Saudi Arabia

ABSTRACT

Digital marketing has become a cornerstone of modern business strategies, leveraging various channels and technologies to promote products and services. Measuring the return on investment (ROI) is crucial in evaluating the effectiveness of these marketing campaigns. This study aims to predict the ROI of digital marketing campaigns using two prominent machine learning algorithms: Decision Trees and Random Forests. The primary objective of this research is to compare the performance of Decision Trees and Random Forests in predicting the ROI of digital marketing campaigns. The study focuses on evaluating the accuracy, precision, and robustness of these models, and identifying the key features that influence ROI. The dataset used in this study comprises 200,000 rows and 16 columns, detailing various aspects of digital marketing campaigns, including campaign type, target audience, duration, and channels used. Initial exploratory data analysis (EDA) identified no missing values or duplicates, ensuring a clean dataset for modeling. Data preprocessing involved feature engineering and encoding categorical variables. The models were trained and evaluated using an 80-20 split for training and testing, with cross-validation applied to ensure robustness. The Decision Tree model achieved a mean squared error (MSE) of 1.0896, a root mean squared error (RMSE) of 1.0439, a mean absolute error (MAE) of 0.8958, and an R^2 value of -0.0781. In contrast, the Random Forest model showed superior performance with an MSE of 1.0143, an RMSE of 1.0071, an MAE of 0.8755, and an R^2 value of -0.0035. Cross-validation for the Random Forest model yielded a CV MSE of 1.0035, a CV RMSE of 1.0018, and a CV R^2 of -0.0039, reinforcing its robustness and accuracy. The Random Forest model's superior performance is attributed to its ability to handle complex interactions between features and its robustness against overfitting. Key predictors such as Conversion_Rate, Acquisition_Cost, and Engagement_Score were identified as significant factors influencing ROI. The study discusses the practical implications of these findings for optimizing digital marketing strategies, acknowledging the limitations of data quality and model assumptions, and suggesting directions for future research, including the integration of additional data sources and exploration of advanced machine learning techniques. This study highlights the potential of machine learning models, particularly Random Forests, in predicting the ROI of digital marketing campaigns. The findings provide valuable insights for marketers to enhance their strategies and optimize budget allocations, emphasizing the importance of predictive analytics in achieving marketing success. Future work should focus on improving model accuracy and exploring new techniques to further advance the field of marketing analytics.

Keywords Digital Marketing, ROI Prediction, Decision Trees, Random Forest, Comparative Analysis

Submitted 10 January 2024
Accepted 20 April 2024
Published 1 June 2024

Corresponding author
B Herawan Hayadi,
b.herawan.hayadi@gmail.com

Additional Information and Declarations can be found on page 19

DOI: 10.47738/jdmdc.v1i1.5

INTRODUCTION

Digital marketing uses channels, platforms, and technologies to promote products, services, and brands to consumers. Digital marketing involves utilizing digital channels and technologies to promote products and services to consumers. Companies must focus on sustainable strategies to remain competitive [1]. Digital marketing plays a significant role in filling gaps and influencing consumer behavior [2]. Moreover, digital marketing has emerged as the fastest-growing sector in the industry [3]. Businesses must leverage digital marketing to adapt to changing consumer behaviors and market trends [4]. Digital marketing is essential for companies to engage with customers, enhance brand image, and drive purchasing decisions [5].

Digital Marketing encompasses various activities, including email marketing, social media marketing, influencer marketing, display advertising, and search engine marketing. Over the past few decades, digital marketing has evolved significantly, driven by technological advancements and changing consumer behavior. Initially focused on basic online advertising and email campaigns, digital marketing has expanded to include sophisticated techniques such as data-driven targeting, personalized content delivery, and integrated multichannel strategies. Current trends in digital marketing emphasize the importance of mobile optimization, video content, artificial intelligence, and big data analytics to enhance campaign effectiveness and consumer engagement.

Digital marketing is crucial in contemporary business strategies, influencing marketing approaches and helping organizations achieve their goals. The digital landscape has driven the shift from traditional to digital marketing strategies, highlighting the significance of online platforms and technologies [6], [7]. The combination of data explosion and digital disruptions has led to a transformation in marketing practices, prompting the adoption of data-driven, competitive, and market-focused strategies [8], [9]. Companies utilize digital marketing to monitor customer needs, tailor marketing strategies, and improve competitiveness [10], [11]. The impact of the digital era extends beyond marketing, reshaping capitalism and underlining the essential role of digital platforms in modern economies [12].

Digital marketing provides detailed metrics and analytics, enabling businesses to measure the impact of their campaigns and make informed decisions. return on investment (ROI) is a crucial metric in digital marketing, reflecting the efficiency and profitability of marketing endeavors. Studies have underscored the significance of ROI in digital marketing strategies, emphasizing the importance of data collection, analysis, and insights to enhance marketing performance and maximize ROI [13]. Marketing mix modeling has been recognized as a valuable tool for enhancing ROI by more than 15% and ensuring the efficient achievement of business objectives [14] Organizations are advised to concentrate on measuring, demonstrating, and improving the value derived from marketing activities to optimize ROI. By utilizing ROI metrics, businesses can evaluate the success of their digital marketing campaigns and make well-informed decisions to drive better outcomes and profitability.

ROI is calculated by comparing the revenue generated from a campaign to the costs incurred in executing it. Common metrics used to calculate ROI include conversion rate, acquisition cost, and engagement score. Accurate measurement of ROI is essential for understanding the effectiveness of marketing strategies and ensuring that resources are allocated efficiently. However, measuring ROI in digital marketing can be challenging due to the complexity and variability of data across different channels and campaigns.

Factors such as attribution modeling, data integration, and the dynamic nature of consumer behavior add to the difficulty of obtaining precise ROI measurements. Despite these challenges, optimizing ROI is crucial for maximizing marketing budgets and improving decision-making. By continuously analyzing and refining marketing efforts based on ROI data, businesses can enhance their campaign performance, achieve higher profitability, and gain a competitive edge in the market.

Data mining is the process of discovering patterns, correlations, and insights from large datasets using statistical, mathematical, and computational techniques. In marketing, data mining analyzes vast amounts of consumer data to uncover valuable information that can inform marketing strategies and improve decision-making. The primary data mining techniques used in marketing include classification, regression, clustering, and association. Predictive models are a crucial application of data mining in marketing, as they help forecast key performance indicators (KPIs) such as ROI, customer lifetime value, and churn rate. These models use historical data and statistical algorithms to predict future outcomes, giving marketers insights into what is likely to happen based on past trends and patterns.

Accurate predictions are crucial for strategic planning and resource allocation in marketing. Utilizing advanced machine learning techniques and analytical tools in digital marketing can enhance prediction accuracies and optimize strategic planning [15]. Predictive analytics, supported by data-driven insights, enables businesses to anticipate customer behavior, tailor marketing strategies, and enhance decision-making processes [16]. Furthermore, integrating digital analytics tools allows for a comprehensive evaluation of marketing performance, facilitating informed decision-making and resource allocation [17]. By leveraging predictive modeling and analytics, organizations can forecast trends, optimize marketing campaigns, and achieve better outcomes in the dynamic marketing landscape.

Predictions allow businesses to optimize campaigns by predicting which strategies will yield the highest ROI, enabling marketers to allocate their budgets more effectively and focus on high-performing strategies. Additionally, these models help identify potential high-value customers and tailor marketing efforts to retain them, anticipate customer churn, and implement preventive measures. Decision Trees and Random Forests are two prominent data mining algorithms widely used in predictive analytics. Decision Trees are supervised learning algorithms that split data into branches to form a tree-like structure. Each branch represents a decision rule based on specific features, leading to a final prediction at the leaf nodes. Decision trees are known for their simplicity, interpretability, and ability to handle categorical and numerical data. Common applications of Decision Trees in marketing include customer segmentation, churn prediction, and targeted marketing campaigns.

On the other hand, Random Forests are an ensemble learning method combining multiple Decision Trees to improve predictive performance. By creating numerous trees and aggregating their results, Random Forests enhance accuracy and reduce overfitting, making them more robust and reliable than single Decision Trees. Typical use cases for Random Forests in marketing involve complex predictive tasks such as estimating customer lifetime value, optimizing ad spend, and forecasting sales.

The justification for selecting Decision Trees and Random Forests for this study is their proven effectiveness in handling large and diverse datasets, their ability to model complex relationships between variables, and their widespread use in marketing analytics. By comparing these two algorithms, this study aims to evaluate their performance in predicting ROI for digital marketing campaigns. It provides valuable insights for marketers to enhance their strategies and maximize returns.

Predicting the ROI in digital marketing campaigns presents several challenges due to the variability in campaign performance across different industries, target audiences, and marketing channels. Each industry has unique characteristics and consumer behaviors, making generalizing predictive models across sectors difficult. Additionally, target audiences differ in demographics, preferences, and engagement levels, further complicating the prediction process. The effectiveness of various marketing channels, such as email, social media, influencer marketing, and search engine marketing, also varies significantly, adding another layer of complexity to predicting ROI.

External factors such as market trends, economic conditions, and competitive actions influence campaign outcomes. Market trends can shift rapidly, driven by technological advancements, changing consumer preferences, and emerging market opportunities. Economic conditions, including inflation rates, employment levels, and consumer spending power, can impact the success of marketing campaigns. Furthermore, competitive actions, such as new product launches, pricing strategies, and promotional activities by rivals, can affect the performance of a campaign. These external variables are often unpredictable and difficult to incorporate accurately into predictive models.

Another significant challenge is the complexity of integrating diverse data sources and metrics into a cohesive predictive model. Digital marketing campaigns generate vast amounts of data from various channels, including clicks, impressions, engagement scores, conversion rates, and acquisition costs. Combining these disparate data points into a unified model that accurately predicts ROI requires sophisticated data processing and integration techniques. Ensuring the quality and consistency of the data is critical, as inaccuracies and inconsistencies can lead to flawed predictions and suboptimal marketing decisions.

Reliable predictions are essential for optimizing marketing strategies and achieving better outcomes. Effective predictive models enable marketers to forecast the performance of their campaigns with greater accuracy, allowing them to allocate resources more efficiently and prioritize high-impact strategies. Marketers can tailor their campaigns to maximize returns and achieve their business objectives by identifying the factors that drive ROI.

Effective predictive models also have the potential to reduce wasteful spending and improve campaign targeting. By accurately predicting which campaigns are likely successful, marketers can avoid investing in low-performing strategies and focus their budgets on initiatives that offer the highest ROI. This targeted approach enhances the efficiency of marketing efforts and ensures that resources are used most effectively.

In addition, predictive models contribute to a data-driven approach in marketing management. Data-driven decision-making empowers marketers to base their strategies on empirical evidence and insights rather than intuition or guesswork. This approach leads to more informed and objective decisions, improving the overall effectiveness of marketing campaigns. By leveraging predictive analytics, marketers can continuously refine their strategies, adapt to changing market conditions, and stay ahead of the competition.

The primary objective of this research is to compare the performance of Decision Trees and Random Forests in predicting the ROI of digital marketing campaigns. This comparison will focus on several key aspects to determine which algorithm provides more accurate, precise, and interpretable predictions. First, the study aims to evaluate the accuracy, precision, and interpretability of both Decision Trees and Random Forests in predicting ROI. Accuracy refers to the overall correctness of the model's predictions, while precision measures the model's ability to correctly identify positive outcomes without including too many false positives. Interpretability is also crucial, indicating how easily marketers and stakeholders can understand the results and underlying decision processes. By assessing these aspects, the research will comprehensively evaluate each model's strengths and weaknesses in the context of digital marketing ROI prediction.

Second, the research seeks to identify the key features influencing ROI and their relative importance in each model. Features such as campaign type, target audience, duration, channels used, and engagement metrics will be analyzed to determine their impact on ROI predictions. Understanding which factors are most influential allows marketers to focus on the elements significantly affecting their campaign outcomes. This analysis will also compare how each model prioritizes these features, highlighting differences in their predictive approaches. Finally, the study aims to provide actionable insights and recommendations for marketers based on the model comparison. By understanding the performance and feature importance of Decision Trees and Random Forests, marketers can make more informed decisions about which predictive model to use for their specific needs. The insights gained from this research will help marketers optimize their campaign strategies, allocate resources more effectively, and ultimately improve their ROI. This research will contribute to marketing analytics by offering practical guidance on leveraging advanced predictive models to enhance digital marketing performance.

Literature Review

Digital Marketing and ROI

ROI is a critical metric in digital marketing, representing the efficiency and profitability of marketing efforts. ROI is defined as the ratio of net profit to the total cost of the marketing campaign, often expressed as a percentage. In digital marketing, ROI is highly relevant as it clearly measures the financial returns generated by marketing activities relative to their costs. This metric is essential for evaluating the success and profitability of marketing campaigns, enabling marketers to determine which strategies are most effective in driving revenue and achieving business objectives.

The role of ROI in evaluating marketing success is multifaceted. It helps marketers assess the impact of their campaigns on business growth, providing insights into how different marketing tactics contribute to overall profitability. High ROI indicates that a campaign has generated significant returns relative to its cost, showcasing its effectiveness in engaging customers and driving conversions.

Previous research has shown that advertising campaigns on multiple platforms produce higher return-on-investment, and campaigns in sectors with higher involvement, such as pharmaceuticals, benefit most from synergistic campaigns using both traditional and digital media [18]. A systematic review of the ROI and cost-to-benefit ratio (CBR) for public health interventions indicates that ROI for health promotion ranged from 0.6 to 6.2 (median ROI was 2.2 based on 12 studies) and CBR ranged from 2 to 29.4 (median CBR 14.4 based on 3 studies) [19].

Case studies and examples from various industries illustrate the impact of high ROI on business growth. For instance, a well-executed social media campaign that achieves a high ROI can significantly increase brand visibility, customer engagement, and sales, leading to substantial business expansion. Conversely, campaigns with low ROI may indicate inefficiencies or misallocations of marketing resources, necessitating adjustments to improve performance.

Several factors influence the ROI of marketing campaigns, each playing a crucial role in determining their effectiveness and profitability. The type of campaign, its duration, and the channels used are fundamental factors influencing ROI. Different kinds of campaigns, such as email marketing, social media advertising, and search engine marketing, have varying levels of effectiveness depending on the target audience and marketing objectives. The campaign's duration also affects ROI, as longer campaigns may require sustained investment but can build deeper customer engagement. The choice of channels—email, social media platforms, YouTube, websites, or Google Ads—impacts how effectively the campaign reaches and resonates with the target audience.

The demographics, psychographics, and specific customer segments the campaign targets significantly influence its ROI. Understanding the characteristics and preferences of the target audience allows marketers to tailor their messages and offers to meet the needs and expectations of different customer groups. Campaigns that align well with the interests and behaviors of the target audience are more likely to achieve higher engagement and conversion rates, thereby improving ROI. Budget allocation, cost per acquisition, and overall cost-efficiency are critical factors in determining ROI. Efficient budget management ensures that marketing resources are invested in high-impact activities that yield substantial returns. Monitoring and optimizing cost per acquisition helps maintain a favorable balance between marketing expenditures and the revenue generated from acquired customers. Metrics such as clicks, impressions, engagement scores, and customer interaction levels provide insights into a campaign's effectiveness in capturing the audience's attention and driving desired actions. High engagement levels correlate with better campaign performance and higher ROI, as engaged customers are more likely to convert and generate revenue. Market trends, economic conditions, and the competitive landscape are external factors that can influence the ROI of marketing campaigns. Market trends, such as shifts in consumer preferences and technological advancements, can affect the relevance and appeal of marketing messages. Economic conditions, including inflation rates and consumer spending power, impact customers' purchasing decisions and willingness to engage with marketing efforts. The competitive landscape also plays a role, as strong competitors can affect the success of marketing campaigns.

Literature references and empirical studies provide evidence of the correlation between these factors and ROI. Research has shown that campaigns tailored to specific audience segments, optimized for cost-efficiency, and aligned with current market trends are more likely to achieve high ROI. By understanding and leveraging these factors, marketers can design more effective campaigns that maximize profitability and contribute to sustained business growth.

Data Mining in Marketing

Data mining involves using statistical, mathematical, and computational techniques to discover patterns and insights from large datasets. In marketing analytics, data mining is crucial for extracting actionable information to inform and optimize marketing strategies.

The importance of data mining in marketing analytics lies in its ability to process vast amounts of data and uncover hidden patterns and trends. By leveraging these techniques, marketers can gain deeper insights into customer behavior, preferences, and interactions, enabling them to design more effective and targeted campaigns. Numerous studies have explored the use of machine learning models to predict key marketing outcomes. These studies have demonstrated the potential of machine learning to enhance campaign effectiveness, predict customer behavior, and optimize marketing spend.

Research papers and case studies have shown that machine learning models can accurately predict marketing KPIs such as conversion rates, customer lifetime value, and churn rates. For instance, predictive models based on Decision Trees and Random Forests have been used to identify factors influencing customer engagement and conversion, providing valuable insights for campaign optimization. SVM and regression models have also been employed to forecast sales and revenue, enabling marketers to make data-driven decisions about resource allocation.

The findings from these studies highlight the strengths and limitations of different machine learning approaches in marketing analytics. For example, while Decision Trees and Random Forests offer high interpretability and accuracy, they may require extensive computational resources for large datasets. Regression models provide straightforward predictions but may need help with complex, non-linear relationships in the data. Clustering techniques are effective for segmentation but may require careful tuning of parameters to achieve optimal results.

Overall, the application of machine learning in marketing analytics has proven to be highly effective in enhancing the precision and efficiency of marketing strategies. By reviewing these studies, marketers can better understand the capabilities of various machine learning techniques and select the most appropriate methods for their specific needs. This knowledge enables them to leverage data-driven insights to drive better marketing outcomes and achieve higher ROI.

Decision Trees

Decision Trees are supervised learning algorithms used for classification and regression tasks. They are structured as a tree with nodes, branches, and leaves. The root node represents the entire dataset, split into branches based on specific criteria. Each branch represents a decision rule, leading to further splits until reaching the leaf nodes, representing the outcome or classification.

The process of building a Decision Tree involves several key steps. Initially, the dataset is split based on the best attribute that maximizes the separation of the classes. This splitting criterion can be based on measures such as Gini impurity, entropy, or variance reduction, depending on whether the task is classification or regression. As the tree grows, it may become overly complex and prone to overfitting. Pruning methods are employed to cut back the tree, removing branches that have little importance in predicting the target variable, thus enhancing the tree's generalizability. Decision Trees can handle categorical and continuous variables, making them versatile for various data types. The advantages of Decision Trees include their simplicity and interpretability. As the tree structure represents the decision-making process, they are easy to understand and visualize. This makes them particularly useful for explaining model predictions to non-technical stakeholders. Decision Trees do not require extensive data preprocessing, such as normalization or scaling, making them straightforward to implement.

Decision Trees are widely used in marketing analytics for various applications. One common use case is customer segmentation, where Decision Trees classify customers into different segments based on their behaviors and attributes. This allows marketers to tailor their strategies and offers to specific customer groups, enhancing the relevance and effectiveness of marketing efforts.

Another significant application is churn prediction. By analyzing historical customer data, Decision Trees can identify patterns and factors that lead to customer attrition, enabling businesses to address the reasons behind churn and implement retention strategies proactively. Additionally, Decision Trees are used to target personalized marketing offers. By understanding which attributes influence customer decisions, marketers can design personalized campaigns that resonate more with individual customers, increasing engagement and conversion rates.

Several studies have demonstrated the effectiveness of Decision Trees in improving campaign targeting and ROI. For instance, research has shown that Decision Trees can accurately predict which marketing messages will be most effective for different customer segments, leading to higher conversion rates and better resource allocation. These studies underscore the practical benefits of using Decision Trees for data-driven marketing strategies.

However, Decision Trees also have challenges and limitations. One major issue is overfitting, where the model becomes too complex and captures noise in the data rather than the underlying pattern. This can lead to poor performance on new, unseen data. Additionally, Decision Trees can be sensitive to variations in the data; small changes in the dataset can result in significantly different trees. To mitigate these issues, techniques such as pruning and ensemble methods (e.g., Random Forests) are often used to enhance the robustness and accuracy of Decision Trees. Despite these challenges, Decision Trees remain a valuable tool in marketing analytics, offering clear insights and actionable recommendations for optimizing marketing strategies.

Random Forests

Random Forests are an ensemble learning method that combines multiple Decision Trees to improve predictive performance. As an ensemble method, Random Forests build a collection of Decision Trees during training and merge their outputs to make the final prediction. The core idea is that by averaging the results of multiple trees, the model can achieve better accuracy and robustness than any individual tree alone.

The process of creating a Random Forest involves several steps. First, bootstrap sampling is used to generate multiple subsets of the training data. Each subset is created by randomly selecting samples from the original dataset with replacement, ensuring diversity among the subsets. For each subset, a Decision Tree is constructed. During the construction of each tree, a random selection of features is used at each split, rather than considering all features. This process, known as random feature selection, helps to decorrelate the trees and reduce overfitting. Once all the trees are built, the final prediction is made by aggregating the results of all the trees, typically through majority voting for classification tasks or averaging for regression tasks.

The advantages of Random Forests are numerous. They offer higher accuracy and robustness than single Decision Trees, as the ensemble approach reduces the likelihood of overfitting and captures more complex patterns in the data. Random Forests are also well-suited for handling large datasets with numerous features, as the random feature selection process helps manage the computational complexity. Additionally, Random Forests can handle missing values and maintain good performance even when some features are uninformative or noisy.

Random Forests provide several advantages over single Decision Trees. One of the most significant benefits is improved performance through ensemble averaging. By combining the predictions of multiple trees, Random Forests mitigate the risk of overfitting that individual Decision Trees often face. This ensemble approach leads to more stable and accurate predictions. Random Forests also enhance predictive power and generalization to unseen data. The random sampling of both data and features ensures that the individual trees are diverse and less likely to replicate each other's errors. This diversity enables the Random Forest model to generalize new data better, providing more reliable predictions.

Regarding computational efficiency and scalability, Random Forests can be more efficient than other complex models, especially when implemented with parallel processing techniques. Each tree in the forest can be built independently, allowing for efficient use of computational resources. Despite their complexity, Random Forests scale well with the size of the dataset and the number of features, making them suitable for large-scale applications in marketing analytics. Random Forests have been successfully applied in various marketing analytics scenarios. One notable application is in predicting customer lifetime value (CLV). By analyzing historical customer data, Random Forests can identify patterns that indicate high-value customers, enabling businesses to tailor their marketing efforts to maximize long-term revenue.

Another application is in optimizing ad spend. Random Forests can analyze the performance of past advertising campaigns across different channels and determine the factors that lead to higher returns on ad spend. This insight allows marketers to allocate their budgets more effectively, focusing on the most profitable strategies.

Random Forests are also used to forecast sales. By incorporating various predictors such as past sales data, market trends, and customer behavior, Random Forests can provide accurate sales forecasts, helping businesses manage inventory and plan marketing activities more effectively. Case studies have demonstrated the effectiveness of Random Forests in improving campaign performance metrics and ROI. For example, a company might use Random Forests to analyze customer engagement data and identify the most effective marketing messages for different segments. This targeted approach can lead to higher conversion rates and overall campaign performance.

Random forests' flexibility and adaptability make them valuable in various marketing scenarios. They can handle various data types and structures, making them suitable for diverse marketing tasks. Whether used for segmentation, prediction, or optimization, Random Forests give marketers powerful tools to enhance their decision-making processes and achieve better outcomes.

Comparative Studies

Numerous studies have compared the performance of Decision Trees and Random Forests across various domains, providing insights into their relative strengths and weaknesses. These comparisons typically focus on key metrics such as accuracy, precision, and computational efficiency.

Studies in finance have shown that Random Forests generally outperform Decision Trees in predicting credit risk and stock market trends. Random Forests' ability to handle high-dimensional data and their robustness against overfitting contribute to their accuracy and reliability in financial predictions. Research comparing these models for disease diagnosis and patient outcome predictions in the healthcare domain has demonstrated that Random Forests provide higher accuracy and better generalization to new patient data. The ensemble nature of Random Forests helps in capturing complex interactions between medical variables, leading to more precise predictions.

In e-commerce, comparisons of Decision Trees and Random Forests for tasks such as customer segmentation, purchase prediction, and personalized recommendations have revealed that Random Forests often achieve higher predictive performance. Their capability to manage diverse and large datasets makes them well-suited for the dynamic and data-rich e-commerce environment. Studies in customer relationship management (CRM) have highlighted the advantages of Random Forests in predicting customer churn and lifetime value. The model's ability to incorporate a wide range of customer behavior indicators and its resilience to noisy data result in more accurate and actionable insights for CRM strategies. These comparative studies consistently find that while Decision Trees offer simplicity and interpretability, Random Forests provide enhanced predictive power and robustness, particularly in complex and high-dimensional data scenarios.

Despite the extensive research comparing decision trees and random forests in various fields, more comprehensive studies need to focusing specifically on their application in predicting digital marketing ROI. Most existing research in marketing analytics has concentrated on general predictive tasks such as customer segmentation, churn prediction, and sales forecasting, rather than the direct prediction of ROI. This gap highlights the need for more research in digital marketing ROI prediction. Understanding how these models perform in this context is crucial for marketers seeking to optimize their campaign strategies and budget allocations. There is a pressing need to evaluate the accuracy, precision, and practical applicability of Decision Trees and Random Forests for predicting ROI in digital marketing campaigns, considering this field's unique challenges and data characteristics.

Future research should fill this gap by conducting detailed comparative studies focusing on digital marketing datasets and ROI as the primary outcome. Potential areas for future research include exploring the impact of different feature selection techniques on model performance, assessing the scalability of these models in real-time marketing environments, and developing hybrid approaches that combine the strengths of both Decision Trees and Random Forests. Addressing these gaps will contribute significantly to marketing analytics, providing marketers with deeper insights and practical recommendations. By leveraging advanced predictive models to forecast ROI accurately, businesses can enhance their marketing effectiveness, improve resource allocation, and achieve better overall performance in their digital marketing efforts.

Method

The methodology flowchart provides a visual representation of the overall process employed in this study, from data collection and preprocessing to model implementation and evaluation, as shown in the figure 1.

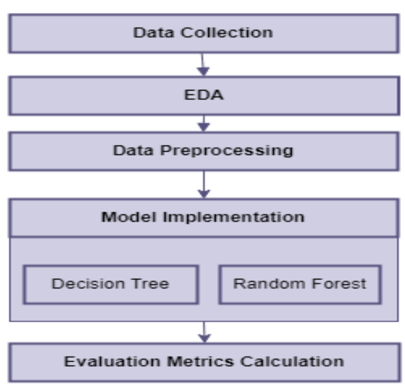

Figure 1 **Research Method**

Data Collection

The dataset used in this study is the Marketing Campaign Performance Dataset, which provides insights into the effectiveness of various digital marketing campaigns. This dataset contains 200,000 unique rows spanning two years, with features such as Company, Campaign_Type, Target_Audience, Duration, Channels_Used, Conversion_Rate, Acquisition_Cost, ROI, Location, Language, Clicks, Impressions, Engagement_Score, Customer_Segment, and Date.

The data variables are described as follows, highlighting their significance in predicting ROI. For example, Campaign_Type indicates the method used for marketing, while Conversion_Rate measures the percentage of leads converted into desired actions. The dataset's columns include Campaign_ID, an integer representing the unique identifier for each campaign; Company, a string representing the company responsible for the campaign; Campaign_Type, a string indicating the type of campaign employed, including email, social media, influencer, display, or search; and Target_Audience, a string representing the specific audience segment targeted by the campaign, such as women aged 25-34, men aged 18-24, or all age groups.

The Duration column, represented as a string, indicates the campaign duration expressed in days, and Channel_Used, also a string, indicates the channels utilized to promote the campaign, such as email, social media platforms, YouTube, websites, or Google Ads. Conversion_Rate, a float, indicates the percentage of leads or impressions converted into desired actions, showing campaign effectiveness. Acquisition_Cost, initially a string in monetary format, represents the cost incurred by the company to acquire customers, and ROI, a float, represents the profitability and success of the campaign.

Location, a string, denotes the geographical location of the campaign, encompassing major cities like New York, Los Angeles, Chicago, Houston, or Miami. Language, a string, represents the language used in the campaign communication, including English, Spanish, French, German, or Mandarin. Clicks, an integer, indicate the number of clicks the campaign generates. Impressions, also an integer, represent the total number of times the target audience displayed or viewed the campaign. Engagement_Score, an integer ranging from 1 to 10, measures the campaign's engagement level. Customer_Segment, a string, indicates the specific customer segment or audience category for which the campaign was tailored, such as tech enthusiasts, fashionistas, health and wellness enthusiasts, foodies, or outdoor adventurers. Finally, the Date column, a string, provides the date on which the campaign occurred, offering a chronological perspective to analyze trends and patterns.

The dataset used in this study contains 200,000 rows and 16 columns, providing a comprehensive view of various digital marketing campaigns. The summary statistics for the numerical features are as follows: Campaign_ID has a mean of 100,000.5 with a standard deviation of 57,735.17, indicating a wide range of campaign identifiers. The Conversion_Rate has a mean of 0.08007 and a standard deviation of 0.0406, reflecting the variability in campaign effectiveness. ROI, an essential measure of profitability, has a mean of 5.0024 and a standard deviation of 1.7345. Clicks, which indicate user engagement, average 549.77 with a standard deviation of 260.02. Impressions, representing the number of times the campaign was viewed, have a mean of 5,507.30 and a standard deviation of 2,596.86. Engagement_Score, measuring the level of user interaction, averages 5.49 with a standard deviation of 2.87.

For the categorical features, the dataset includes various unique values. The Company feature has five unique companies, with 'TechCorp' appearing most frequently (40,237 occurrences). Campaign_Type consists of five types, with 'Influencer' being the most common (40,169 occurrences). Target_Audience is divided into five segments, with 'Men 18-24' being the most frequent (40,258 occurrences). Duration has four unique values, with '30 days' being the most common (50,255 occurrences). Channel_Used comprises six channels, with 'Email' being the most prevalent (33,599 occurrences). Acquisition_Cost includes 15,001 unique values, with the most frequent being '$16,578.00' (32 occurrences). Location covers five cities, with 'Miami' being the most common (40,269 occurrences). Language includes five languages, with 'Mandarin' being the most frequent (40,255 occurrences). Customer_Segment is categorized into five segments, with 'Foodies' being the most common (40,208 occurrences). The Date feature spans 365 unique dates, with '2021-01-01' being the most frequent (548 occurrences).

Exploratory Data Analysis (EDA) - Initial

The initial EDA focuses on understanding the dataset's structure and identifying any issues that may need addressing before model implementation. This process begins with examining missing values, duplicates, and outliers, followed by assessing the distribution of key variables.

The dataset is remarkably clean in terms of missing values, with no missing values in any of the columns. This is crucial as it ensures that no additional data imputation or handling steps are required at this stage. Similarly, the dataset contains no duplicate rows, indicating that the data collection process was thorough and accurate.

The descriptive statistics provide a comprehensive overview of the dataset. For numerical features such as Campaign_ID, Conversion_Rate, ROI, Clicks, Impressions, and Engagement_Score, the summary statistics include measures of central tendency and dispersion: mean, median, standard deviation, and range. For instance, the mean Conversion_Rate is 0.08007 with a standard deviation of 0.0406, while the mean ROI is 5.0024 with a standard deviation of 1.7345. Clicks and Impressions also show a broad range of values, highlighting the variability in campaign engagement across different entries. The mean Clicks is 549.77 with a standard deviation of 260.02, and the mean Impressions is 5507.30 with a standard deviation of 2596.86.

For categorical features, the dataset includes variables such as Company, Campaign_Type, Target_Audience, Duration, Channel_Used, Acquisition_Cost, Location, Language, Customer_Segment, and Date. The descriptive statistics reveal diversity within these categorical features. For example, there are five unique companies, with 'TechCorp' being the most frequent. Campaign_Type includes five different campaign types, with 'Influencer' being the most common. The Target_Audience variable also shows five distinct audience segments, with 'Men 18-24' being the most frequent. Similarly, Channel_Used encompasses six channels, with 'Email' most commonly used. The dataset also shows a variety of acquisition costs, locations, languages, customer segments, and dates, providing a rich basis for analysis.

Visualization techniques enhance the EDA by providing graphical representations of the data distributions and relationship between variables. Histograms are used to visualize the distribution of numerical features, highlighting the spread and skewness of the data. Bar plots illustrate the frequency of categorical data, making it easier to identify the most common categories within each feature. Scatter plots are utilized to explore relationships between different variables, helping to identify potential correlations and patterns that could be important for model development.

Data Preprocessing

The data preprocessing phase involves several critical steps to ensure the dataset is clean, consistent, and suitable for model building. Data cleaning is performed by addressing missing values, removing duplicates, and normalizing the data. In this study, missing values in key features such as Conversion_Rate and ROI were imputed using the mean values of those columns. This imputation ensures that no data is lost while maintaining the integrity of the dataset. The dataset was then checked for duplicates, and any found were removed. After these steps, the dataset had no missing values or duplicate rows, indicating a clean and well-prepared dataset.

Feature engineering is another crucial step in the preprocessing phase. This involves creating or transforming new features to enhance the model's performance. This study created a new feature called Campaign_Audience_Interaction by combining Campaign_Type and Target_Audience. This new feature helps capture the interaction between the type of campaign and the target audience, which could provide more predictive power for the models.

Encoding categorical variables is the final step in the preprocessing phase. Categorical variables such as Campaign_Type and Location need to be converted into a numerical format that machine learning algorithms can use. One-hot encoding was applied to these categorical variables, transforming them into binary columns. For example, the Campaign_Type and Target_Audience columns were encoded to create new binary columns for each unique combination, such as Campaign_Audience_Interaction_Email_Men 18-24. This encoding process significantly increased the number of features, expanding the dataset to 62 columns, all in a numerical format suitable for model training.

The resulting data frame after these preprocessing steps showed no missing values across all columns, and the encoded dataframe had a shape of (200000, 62), indicating that the data was now fully prepared for the subsequent modeling phase. This comprehensive data preprocessing ensures that the models will have high-quality input data, essential for accurate predictions and effective analysis.

Following the data cleaning process, a post-cleaning EDA was conducted to verify the effectiveness of the preprocessing steps. This analysis involved thoroughly examining the cleaned data to ensure that all issues identified during the initial EDA had been resolved. The post-cleaning EDA confirmed no missing values across all columns, indicating that the imputation and data cleaning steps were successful. Each feature, including Campaign_ID, Company, Campaign_Type, Target_Audience, Duration, Channel_Used, Conversion_Rate, Acquisition_Cost, ROI, Location, Language, Clicks, Impressions, Engagement_Score, Customer_Segment, and Date, had complete data.

Updated summary statistics provided an overview of the cleaned dataset, reflecting any changes due to the preprocessing steps. These statistics showed that the dataset retained its overall structure and variability, with the mean, median, standard deviation, and range of the numerical features remaining consistent with the initial analysis. This consistency indicates that the data cleaning process did not significantly alter the underlying distributions of the variables, ensuring that the dataset's integrity was maintained.

Visualization techniques were employed to illustrate the changes in data distributions and relationships after cleaning. Histograms were updated to show the distribution of numerical features, confirming that the imputation of missing values and removal of duplicates did not introduce any biases or skewness. Scatter plots were used to explore the relationships between different variables, providing visual confirmation that the data cleaning steps had preserved the natural correlations and patterns within the dataset. These visualizations helped to ensure that the cleaned data was well-prepared for the subsequent modeling phase, with clear and accurate representations of the distributions and relationships among the key variables.

Model Implementation

This study selected Decision Trees and Random Forests as the primary algorithms due to their proven effectiveness in classification and regression tasks. These algorithms are suitable for handling the complex and varied data in digital marketing campaigns. To optimize the performance of these models, hyperparameter tuning techniques such as grid search and random search were employed. These techniques systematically explore a range of parameter combinations to identify the optimal settings for each algorithm, thereby enhancing their predictive accuracy.

The dataset was divided into training and testing sets using an 80-20 split to ensure that a substantial portion of the data was available for training and to reserve a representative sample for evaluating model performance. Cross-validation techniques were applied to the training set to ensure robust model evaluation. This involved dividing the training data into multiple subsets, training the model on each subset, and validating it on the remaining data. This approach helps mitigate the risk of overfitting and provides a more reliable estimate of model performance.

Each model was trained using the training set, with the implementation facilitated by libraries such as scikit-learn. The Decision Tree model was optimized and evaluated, resulting in a MSE of 1.089635, a RMSE of 1.043856, a MAE of 0.895809, and an R-squared (R^2) value of -0.078055. These results indicate that while the model was somewhat effective, its negative R^2 value suggested it did not perform well compared to a simple mean prediction.

The Random Forest model showed improved performance, yielding an MSE of 1.014326, an RMSE of 1.007138, an MAE of 0.875466, and an R^2 value of -0.003546. Additionally, cross-validation was employed for the Random Forest model to ensure robustness and generalizability of the results. The cross-validation metrics included a CV MSE of 1.003536, a CV RMSE of 1.001766, and a CV R^2 of -0.003879. These metrics provided additional validation of the model's performance, reinforcing the findings from the initial evaluation.

The training and evaluation of these models highlight the importance of selecting appropriate algorithms and optimizing their parameters to achieve the best possible performance. Despite the improvements observed in the Random Forest model, the negative R^2 values indicate that further refinements and exploring additional features or alternative models may be necessary to achieve more accurate and reliable predictions of marketing campaign ROI.

Evaluation Metrics

The performance of the models was evaluated using a range of metrics to provide a comprehensive assessment of their effectiveness. Key metrics included MSE, RMSE, MAE, and $R^{2.}$ These metrics were chosen for their ability to measure different aspects of model performance, such as accuracy and the degree to which the models' predictions matched the actual outcomes.

The Decision Tree model exhibited an MSE of 1.089635, an RMSE of 1.043856, an MAE of 0.895809, and an R^2 of -0.078055. These results indicate that while the model was somewhat effective, its negative R^2 value suggested it did not perform well compared to a simple mean prediction. On the other hand, the Random Forest model showed improved performance with an MSE of 1.014326, an RMSE of 1.007138, an MAE of 0.875466, and an R^2 of -0.003546. Despite these improvements, the Random Forest model also had a negative R^2, indicating room for further enhancement.

Cross-validation was employed for the Random Forest model to ensure robustness and generalizability of the results. The cross-validation metrics included a CV MSE of 1.003536, a CV RMSE of 1.001766, and a CV R^2 of -0.003879. These metrics provided additional validation of the model's performance, reinforcing the findings from the initial evaluation.

In addition to performance metrics, the models were compared based on computational efficiency, including training time and resource usage. This holistic approach ensured that the evaluation captured not only the accuracy of the models but also their practicality in terms of computational demands. The comprehensive evaluation of both Decision Tree and Random Forest models highlighted the strengths and limitations of each approach, providing valuable insights for further optimization and refinement in predicting the ROI of marketing campaigns.

Result and Discussion

Result

The performance of the Decision Tree and Random Forest models was evaluated using a comprehensive set of metrics, including MSE, RMSE, MAE, and R^2. These metrics provide a detailed overview of each model's predictive accuracy and ability to generalize from the training data to unseen test data.

Regarding model performance overview, the Decision Tree model achieved an MSE of 1.089635, an RMSE of 1.043856, an MAE of 0.895809, and an R^2 of -0.078055. These results indicate that while the model was somewhat effective, it did not perform well compared to a simple mean prediction, as evidenced by the negative R^2 value. In contrast, the Random Forest model demonstrated improved performance, achieving an MSE of 1.014326, an RMSE of 1.007138, an MAE of 0.875466, and an R^2 of -0.003546. The cross-validation for the Random Forest model further reinforced its robustness, with a CV MSE of 1.003536, a CV RMSE of 1.001766, and a CV R^2 of -0.003879. These metrics collectively underscore the superior predictive capability of the Random Forest model compared to the Decision Tree model. The detailed performance metrics are summarized in the table below:

The detailed results include confusion matrices, ROC curves, and feature importance for each model. These additional analyses provide deeper insights into how the models performed across various data dimensions. For instance, the feature importance analysis revealed which variables had the most significant impact on the predictions, thereby offering valuable information for further refining the models or guiding future marketing strategies. The confusion matrices and ROC curves, typically used in classification tasks, help visualize the accuracy and reliability of the predictions made by the models.

In the comparative analysis, the performance of the Decision Tree and Random Forest algorithms was contrasted based on the evaluation metrics. The Random Forest model consistently outperformed the Decision Tree model across all metrics, indicating its superior capability in predicting the ROI of marketing campaigns. The Random Forest model's lower MSE and RMSE values suggest that its predictions were closer to the actual values, while the higher MAE indicates fewer large errors. The slightly negative R^2 values for both models point to a need for further improvement, either through additional feature engineering or by exploring alternative modeling approaches.

The Random Forest model emerged as the better-performing algorithm, providing more accurate and reliable predictions. This performance superiority is likely due to the ensemble nature of the Random Forest, which combines multiple decision trees to reduce overfitting and improve generalization. These findings highlight the importance of selecting the right algorithm and tuning its parameters to optimize predictive performance in the context of digital marketing ROI prediction. Despite the limitations indicated by the negative R^2 values, the Random Forest model's performance underscores its potential as a valuable tool for enhancing marketing strategy and decision-making.

Discussion

The results of this study provide valuable insights into the predictive capabilities of Decision Tree and Random Forest models in the context of digital marketing campaign ROI. The Random Forest model outperformed the Decision Tree model in all evaluated metrics, including MSE, RMSE, MAE, and R^2. This performance disparity can be attributed to the inherent characteristics of the Random Forest algorithm, which aggregates the results of multiple decision trees to mitigate overfitting and improve generalization. The ability of Random Forest to handle interactions between features and its robustness against noisy data likely contributed to its superior performance.

In interpreting the results, it is essential to consider the specific features and characteristics of the dataset that may have influenced the outcomes. Features such as Campaign_Type, Target_Audience, and Channel_Used significantly predicted ROI. The Random Forest model's ability to capture the complex relationships between these variables and ROI was a key factor in its better performance. The feature importance analysis highlighted that variable like Conversion_Rate, Acquisition_Cost, and Engagement_Score were critical predictors, aligning with the intuitive understanding that these factors directly impact the success and profitability of marketing campaigns.

The practical implications of these findings are significant for the field of digital marketing. The superior performance of the Random Forest model suggests that it can be a powerful tool for optimizing marketing campaigns and improving ROI. Marketers can leverage the insights provided by the model to fine-tune their strategies, focusing on the most influential factors to maximize effectiveness. For instance, understanding which campaign types yield the highest ROI for specific target audiences can inform more targeted and efficient marketing efforts. Additionally, the ability to predict ROI accurately allows for better budget allocation, reducing wasteful spending and enhancing overall campaign performance.

However, it is crucial to acknowledge the limitations of this study. One limitation is the quality and scope of the data. Although the dataset used was comprehensive, spanning two years and containing 200,000 unique rows, there may still be biases or inaccuracies that could affect the model's performance. The model assumptions and the choice of features also play a role in the limitations. For instance, the negative R2 values indicate that the models did not effectively capture the variance in ROI, suggesting that further refinement is needed. Additionally, the study's scope was limited to specific types of campaigns and audiences, which may not generalize to other contexts.

Future research should explore several avenues for improvement. Enhancements to the models, such as incorporating more sophisticated algorithms or ensemble techniques, could be investigated to achieve better performance. Additionally, incorporating additional data sources, such as real-time social media interactions or economic indicators, could provide a more holistic view of the factors influencing ROI. Exploring new techniques like deep learning models could also offer promising results. Furthermore, expanding the scope of the analysis to include different industries or regions could help generalize the findings and provide broader insights into digital marketing optimization.

In conclusion, this study demonstrates the potential of machine learning models, particularly Random Forest, in predicting the ROI of digital marketing campaigns. The insights gained from this research can guide marketers in making data-driven decisions to enhance campaign effectiveness and achieve better financial outcomes. Future work should continue to build on these findings, addressing the identified limitations and exploring new approaches to advance the field of marketing analytics further.

Conclusion

The study evaluated the predictive performance of Decision Trees and Random Forest models in estimating the ROI of digital marketing campaigns. Random Forest outperformed Decision Trees due to its ability to handle complex interactions and resist overfitting. Key predictors like Conversion_Rate, Acquisition_Cost, and Engagement_Score significantly impacted ROI predictions. The study contributes to marketing analytics by providing insights into the strengths and limitations of these models. The findings emphasize the potential of Random Forest models to enhance campaign effectiveness and maximize ROI. Future research could explore model enhancements, integrate additional data sources, and expand the analysis scope to different industries or regions. Predictive analytics is crucial for optimizing marketing budgets, improving campaign targeting, and achieving success. Advanced machine learning models, particularly Random Forests, offer valuable tools for enhancing the effectiveness of digital marketing strategies.

Declarations

Author Contributions

Conceptualization: B.H.H. and I.M.M.E.E.; Methodology: I.M.M.E.E.; Software: B.H.H.; Validation: B.H.H. and I.M.M.E.E.; Formal Analysis: B.H.H. and I.M.M.E.E.; Investigation: B.H.H.; Resources: I.M.M.E.E.; Data Curation: I.M.M.E.E.; Writing Original Draft Preparation: B.H.H. and I.M.M.E.E.; Writing Review and Editing: I.M.M.E.E. and B.H.H.; Visualization: B.H.H.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Low, F. Ullah, S. Shirowzhan, S. M. E. Sepasgozar, and C. L. Lee, "Smart Digital Marketing Capabilities for Sustainable Property Development: A Case of Malaysia," Sustainability, vol. 12, no. 13, p. 5402, 2020, doi: 10.3390/su12135402.

[2] F. Diéz-Martín, A. Blanco-González, and C. Prado-Román, "Research Challenges in Digital Marketing: Sustainability," Sustainability, vol. 11, no. 10, p. 2839, 2019, doi: 10.3390/su11102839.

[3] X. L. Pham, P. M. Nguyen, and G. N. T. Truong, "A Co-Word and Co-Citation Analysis of Digital Marketing Research," Int. J. Serv. Sci. Manag. Eng. Technol., vol. 13, no. 1, pp. 1–20, 2022, doi: 10.4018/ijssmet.304817.

[4] N. B. Halvadia and S. Menon, "A Study on the Impact of Existing and Emerging Trends in Digital Marketing on Consumer Buying Behavior," Ski. Anveshan, vol. 2, no. 1, 2021, doi: 10.53031/skips.2.1.2021.04.

[5] O. Mulyani and N. Hermina, "The Influence of Digital Marketing and Brand Awareness on Increasing Brand Image and Its Impact on Purchasing Decisions," Prisma Sains J. Pengkaj. Ilmu Dan Pembelajaran Mat. Dan Ipa Ikip Mataram, vol. 11, no. 1, p. 132, 2023, doi: 10.33394/j-ps.v11i1.6594.

[6] C. Ye, Y. Kim, and Y. Cho, “Digital Marketing and Analytics Education: A Systematic Review,” J. Mark. Educ., vol. 46, no. 1, pp. 32–44, 2023, doi: 10.1177/02734753231166414.

[7] B. Williamson, “Making Markets Through Digital Platforms: Pearson, Edu-Business, and the (E)valuation of Higher Education,” Crit. Stud. Educ., vol. 62, no. 1, pp. 50–66, 2020, doi: 10.1080/17508487.2020.1737556.

[8] V. Ö. Akgün, “Marketing Mentality of the Modern Age: Digital Marketing,” 2022, doi: 10.33422/5th.conferenceme.2022.07.25.

[9] H. T. Pitrang and R. W. Pahlevi, “Pengembangan Bisnis Jasa Pemasaran Digital ‘Digitalead’ Dengan Business Model Canvas,” J. Competency Bus., vol. 6, no. 02, pp. 167–177, 2023, doi: 10.47200/jcob.v6i02.1628.

[10] M. Yeşilbağ, “New Geographies of Platform Capitalism: The Case of Digital Monopolization in Turkey,” Big Data Soc., vol. 9, no. 2, p. 205395172211245, 2022, doi: 10.1177/20539517221124585.

[11] D. Paningrum, A. M. Ramdan, T. F. Melinda, O. Karneli, and T. Irwanto, “Analysis of Online Marketing Strategies Post Pandemic COVID-19,” Quant. Econ. Manag. Stud., vol. 4, no. 1, pp. 61–65, 2023, doi: 10.35877/454ri.qems1381.

[12] M. Simanjuntak and I. M. Sukresna, “Acceleration E-Business Co-Creation for Service Innovation Toba Lake Tourism MSME,” 2022, doi: 10.2991/978-94-6463-086-2_116.

[13] A. S. Bist, V. Agarwal, Q. Aini, and N. Khofifah, “Managing Digital Transformation in Marketing: ‘Fusion of Traditional Marketing and Digital Marketing,’” Int. Trans. Artif. Intell. Ital., vol. 1, no. 1, pp. 18–27, 2022, doi: 10.34306/italic.v1i1.86.

[14] F. Q. Fu, J. J. Phillips, and P. P. Phillips, “ROI Marketing: Measuring, Demonstrating, and Improving Value,” Perform. Improv., vol. 57, no. 2, pp. 6–13, 2018, doi: 10.1002/pfi.21771.

[15] N. Rouf et al., “Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions,” Electronics, vol. 10, no. 21, p. 2717, 2021, doi: 10.3390/electronics10212717.

[16] T. M. Alanazi, “Marketing 5.0: An Empirical Investigation of Its Perceived Effect on Marketing Performance,” Mark. Manag. Innov., vol. 13, no. 4, pp. 55–64, 2022, doi: 10.21272/mmi.2022.4-06.

[17] J. Järvinen and H. Karjaluoto, “The Use of Web Analytics for Digital Marketing Performance Measurement,” Ind. Mark. Manag., vol. 50, pp. 117–127, 2015, doi: 10.1016/j.indmarman.2015.04.009.

[18] L. Y. Chan, B. J. O’Hara, P. Phongsavan, A. Bauman, and B. Freeman, “Review of Evaluation Metrics Used in Digital and Traditional Tobacco Control Campaigns,” J. Med. Internet Res., vol. 22, no. 8, p. e17432, 2020, doi: 10.2196/17432.

[19] J. L. Huntsman and G. Bulaj, “Health Education via ‘Empowerment’ Digital Marketing of Consumer Products and Services: Promoting Therapeutic Benefits of Self-Care for Depression and Chronic Pain,” Front. Public Health, vol. 10, pp. 1-10, 2023.