Predicting Customer Conversion in Digital Marketing: Analyzing the Impact of Engagement Metrics Using Logistic Regression, Decision Trees, and Random Forests

Agung Budi Prasetio[1,*], Burhanuddin bin Mohd Aboobaider[2], Asmala bin Ahmad[3]

[1]Faculty Computer Science, Institut Teknologi Tangerang Selatan, Komplek Komersial BSD Kav 9, Jl. Raya Serpong, Lengkong Karya, Serpong Utara, Kota Tangerang Selatan 12246, Indonesia

[2,3]Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Durian Tunggal 76100, Melaka, Malaysia

ABSTRACT

This research explores the impact of engagement metrics on predicting customer conversion rates within digital marketing, employing three advanced predictive modeling techniques: Logistic Regression, Decision Trees, and Random Forests. Using a comprehensive dataset of 8,000 customer interactions, the study evaluates critical engagement metrics such as PagesPerVisit, TimeOnSite, and EmailClicks to determine their influence on conversion outcomes. The results indicate that PagesPerVisit and TimeOnSite are the most significant predictors of customer conversion, with the Random Forest model outperforming others, achieving an accuracy of 87.1% and an ROC-AUC score of 0.6979. The Logistic Regression model demonstrated the highest recall for the conversion class at 99.8%, but its performance in predicting non-conversions was less robust, highlighting the challenges of imbalanced datasets. Decision Trees, while offering valuable interpretability, showed a lower accuracy of 79.6% and struggled with precision in identifying non-conversions. These findings suggest that enhancing on-site customer engagement and refining email marketing strategies are pivotal for improving conversion rates. The study contributes to the field of digital marketing analytics by providing empirical evidence on the relative importance of various engagement metrics and offering practical insights for optimizing digital marketing strategies. Additionally, it highlights the benefits of using ensemble methods like Random Forests to achieve more balanced and accurate predictions in customer conversion scenarios.

Keywords Digital Marketing, Customer Conversion, Engagement Metrics, Predictive Modeling, Random Forests

Submitted 24 January 2025
Accepted 20 April 2025
Published 1 June 2025

Corresponding author
Agung Budi Prasetio,
agung@itts.ac.id

Additional Information and Declarations can be found on page 202

DOI: 10.47738/jdmdc.v2i2.34

© Copyright 2025 Prasetio, et al.,

Distributed under Creative Commons CC-BY 4.0

INTRODUCTION

Digital marketing has become an integral part of the modern business landscape, fundamentally transforming how companies connect with their customers. In today's highly competitive and digitally-driven environment, businesses of all sizes leverage digital marketing strategies to reach a wider audience, engage potential customers, and drive sales. The proliferation of internet access and the widespread use of smartphones have exponentially increased the potential touchpoints for customer interactions, making digital

How to cite this article: A. B. Prasetio, B. M. Aboobaider, A. b. Ahmad, "Predicting Customer Conversion in Digital Marketing: Analyzing the Impact of Engagement Metrics Using Logistic Regression, Decision Trees, and Random Forests," *J. Digit. Mark. Digit. Curr.*, vol. 2, no. 2, pp. 181-204, 2025.

marketing not just an option but a necessity for business growth and sustainability.

The advent of advanced technologies and data analytics has further revolutionized digital marketing, enabling businesses to gather and analyze vast amounts of data on customer behavior and preferences. This data-driven approach allows for highly targeted marketing campaigns, personalized customer experiences, and real-time adjustments to marketing strategies. Consequently, digital marketing has evolved from traditional, one-size-fits-all approaches to highly sophisticated, dynamic, and interactive methods that maximize customer engagement and conversion rates.

In the dynamic landscape of digital marketing, customer engagement has become a pivotal factor in the success of marketing campaigns. Metrics such as website visits, pages per visit, time spent on site, social shares, email opens, and email clicks offer valuable insights into how customers interact with digital content. These metrics not only serve as indicators of customer interest and involvement but also provide marketers with critical data to evaluate the effectiveness of their strategies and enhance the overall customer experience. Understanding and optimizing these engagement metrics is essential for businesses aiming to increase conversion rates, where conversion is defined as the successful completion of a desired action, such as a purchase or newsletter sign-up. High levels of engagement are often correlated with higher conversion rates, suggesting that customers who engage more frequently and deeply with digital content are more likely to convert. Thus, analyzing the impact of these metrics on conversion rates can yield actionable insights, enabling businesses to refine their marketing strategies and allocate resources more efficiently.

Customer engagement plays a fundamental role in building loyalty and driving the success of marketing campaigns. Meaningful interactions and positive experiences are crucial for fostering customer loyalty and satisfaction, as demonstrated by studies indicating that emotional connections to campaign messages can lead to increased customer loyalty [1], [2]. Additionally, engagement acts as a mediator, strengthening the relationship between marketing initiatives and customer loyalty [3]. While research has shown that social media marketing activities have a positive but relatively weak correlation with customer engagement, the interactive nature of social media platforms still makes them vital for influencing consumer purchase behavior [4], [5]. Furthermore, the integration of advanced technologies like AI in marketing strategies is transforming customer engagement, offering enhanced targeting, personalization, and campaign performance [6].

Despite the recognized importance of engagement metrics, there remains a need for a deeper understanding of how these specific metrics influence conversion rates. While previous research has highlighted the general relationship between customer engagement and conversion, there is a gap in the literature regarding the individual contributions of various engagement metrics to conversion outcomes. This study aims to fill that gap by conducting a detailed analysis of how different engagement metrics—namely website visits, pages per visit, time on site, social shares, email opens, and email clicks—impact conversion rates in digital marketing campaigns.

By employing advanced predictive modeling techniques such as logistic regression, decision trees, and random forests, this research seeks to identify which engagement metrics are the strongest predictors of conversion. The insights gained from this analysis will not only enhance our understanding of customer behavior in the digital space but also provide practical guidance for marketers looking to optimize their campaigns for better conversion outcomes. This focused investigation into the nuances of engagement metrics and their relationship with conversion will contribute to the ongoing development of more effective, data-driven digital marketing strategies.

The rapid growth and innovation in digital marketing over the past decade have fundamentally reshaped how businesses connect with their target audiences. As companies increasingly rely on digital channels to engage customers, the effectiveness of these efforts is often gauged through various engagement metrics that track interactions with marketing content. However, the critical relationship between these engagement metrics and the ultimate goal of conversion—whether it involves making a purchase, signing up for a service, or completing another desired action—remains an essential area for ongoing research. Understanding how these metrics influence conversion can provide valuable insights for optimizing marketing strategies, leading to more effective campaigns and improved business outcomes.

Technological advancements and evolving consumer behaviors have driven the evolution of digital marketing, resulting in the adoption of innovative strategies that enhance customer engagement and drive sales growth. The integration of advanced technologies like data analytics and artificial intelligence (AI) has transformed businesses' ability to understand consumer behavior and preferences, enabling the creation of personalized marketing campaigns that resonate more deeply with target audiences. This personalization not only improves customer satisfaction and loyalty but also helps businesses differentiate themselves in a crowded marketplace, ultimately leading to better customer retention and advocacy [7], [8]. Furthermore, the rise of growth hacking as a strategy within digital marketing exemplifies how companies like Uber have achieved exponential growth through rapid experimentation and innovative marketing tactics [9]. This approach, along with the adaptability required during the COVID-19 pandemic, highlights the importance of innovation and resilience in digital marketing, particularly for small and medium enterprises (SMEs) that have leveraged these strategies to compete with larger firms and improve market performance [10], [11], [12].

The primary objective of this research is to analyze the impact of different engagement metrics on conversion rates. This involves a detailed examination of several key metrics, including website visits, pages per visit, time on site, social shares, email opens, and email clicks. Each of these metrics provides a different perspective on customer behavior and engagement, offering potential indicators of their likelihood to convert. By analyzing these metrics, the study aims to identify patterns and correlations that can help marketers better understand which aspects of customer engagement are most influential in driving conversions.

Furthermore, this research seeks to determine which engagement metrics serve as the strongest predictors of conversion. While it is known that higher engagement generally correlates with higher conversion rates, the specific contribution of each metric remains unclear. Using advanced predictive modeling techniques such as logistic regression, decision trees, and random forests, this study will assess the predictive power of each engagement metric. Logistic regression will help quantify the relationship between metrics and conversion probability, decision trees will illustrate decision-making pathways, and random forests will enhance prediction accuracy by aggregating multiple decision trees.

By achieving these objectives, the research aims to provide actionable insights that can be used to optimize digital marketing campaigns. Marketers will be able to focus their efforts on the most impactful engagement metrics, allocate resources more efficiently, and develop strategies that are more likely to convert engaged customers into actual customers. The findings will contribute to the existing body of knowledge in digital marketing analytics and offer practical recommendations for enhancing the effectiveness of digital marketing efforts.

Understanding the factors that drive customer conversions in digital marketing is of paramount importance for marketers seeking to optimize their strategies. This study provides practical insights into how various engagement metrics influence conversion rates, enabling marketers to refine their approaches and achieve better results. By identifying which engagement metrics are the most significant predictors of conversion, marketers can tailor their campaigns to focus on the activities that are most likely to lead to successful outcomes. This targeted approach can lead to more efficient allocation of marketing resources, higher return on investment (ROI), and ultimately, greater business success.

For instance, if the study finds that metrics like email opens and website visits are strong predictors of conversion, marketers can prioritize these areas in their campaigns. They might invest in more compelling email content and strategies to drive website traffic, knowing that these efforts are likely to result in higher conversion rates. Such data-driven strategies not only enhance the effectiveness of marketing campaigns but also contribute to a more personalized customer experience, fostering stronger customer relationships and loyalty.

In addition to its practical applications, this research contributes to the existing body of knowledge in digital marketing analytics. While there is substantial research on customer engagement and conversion, the specific impact of individual engagement metrics has not been thoroughly explored. This study fills that gap by providing a detailed analysis of how each metric affects conversion rates, offering new insights that can inform future research and practice. By employing advanced predictive modeling techniques such as logistic regression, decision trees, and random forests, the study also demonstrates the value of these methods in digital marketing analytics.

The findings of this research have the potential to influence the academic field of digital marketing by providing a robust framework for analyzing engagement metrics and their impact on conversion. This framework can be used by other researchers to further investigate related topics, leading to a deeper understanding of digital marketing dynamics. Additionally, the use of predictive modeling techniques in this study showcases their applicability in marketing analytics, encouraging their adoption in both academic and practical settings. Through these contributions, the study not only enhances the current understanding of digital marketing strategies but also paves the way for more effective and efficient marketing practices in the future.

Literature Review

Overview of Digital Marketing Concepts

Digital marketing encompasses a broad range of strategies and tactics that businesses employ to connect with customers through digital channels. This includes methods such as search engine optimization (SEO), pay-per-click (PPC) advertising, social media marketing, content marketing, email marketing, and affiliate marketing. Each channel offers unique opportunities to reach specific target audiences, deliver personalized messages, and drive customer actions. The advent of the internet and digital technologies has significantly enhanced marketers' ability to collect vast amounts of data on customer interactions, allowing for more precise targeting and personalization than traditional marketing methods. This shift has transformed digital marketing into a data-driven discipline where decisions are increasingly based on empirical evidence rather than intuition alone [13].

A key component of digital marketing is the ability to track and measure various metrics that provide insights into how customers interact with marketing efforts. Metrics such as impressions, clicks, click-through rates (CTR), conversion rates, and bounce rates are commonly monitored to gauge the effectiveness of different marketing strategies. The data collected from these metrics helps marketers understand the customer journey, identify areas for improvement, and optimize campaigns to achieve better results. The integration of data analytics into digital marketing has thus revolutionized the field, allowing organizations to refine their strategies continuously and achieve more targeted, efficient, and effective outcomes [14]. Moreover, the incorporation of advanced technologies like artificial intelligence (AI) and the Internet of Things (IoT) is revolutionizing digital marketing practices, enhancing targeting, personalization, and overall campaign performance [15], [16]. This evolution of digital marketing into a highly sophisticated and data-driven discipline underscores its critical role in modern business strategies [17].

Importance of Customer Engagement and Conversion Metrics

Customer engagement, which refers to the level of interaction and involvement a customer has with a brand's digital presence, plays a crucial role in building strong relationships between customers and brands. High engagement typically signals that customers find the content valuable and are more likely to develop loyalty to the brand. Key engagement metrics such as website visits, pages per visit, time on site, social media interactions, email opens, and clicks provide valuable insights into customer behavior, offering marketers a clearer understanding of what captures customer interest and how they interact with digital content. These metrics are essential for refining marketing strategies to better meet customer needs. On the other hand, conversion metrics measure the success of turning engaged customers into paying customers or achieving other desired actions, such as completing a purchase or signing up for a newsletter. The conversion rate, which reflects the percentage of engaged customers who complete the desired action, is a critical indicator of the effectiveness of marketing efforts in driving business goals. Understanding the relationship between engagement metrics and conversion is vital for optimizing digital marketing strategies, as research has shown that higher engagement often correlates with higher conversion rates, though this relationship can vary across different contexts and industries.

Moreover, customer engagement and conversion metrics are closely intertwined and essential components of successful digital marketing strategies. Engagement not only drives emotional connections and loyalty but also acts as a mediator that enhances the relationship between marketing efforts and conversion outcomes. For example, research has shown that elements like emojis in digital marketing communications can evoke positive emotions, leading to increased customer engagement and purchase intent [18]. Highly engaged customers are more likely to generate content, refer products, and co-create customer experiences, which can influence brand advocacy and positive electronic word-of-mouth [19]. Additionally, strategies that emphasize interactivity and effective site design have been proven to boost business performance by enhancing customer engagement and driving online transactions [20]. Personalized content delivery and data-driven decision-making are also critical in improving customer experience and driving business growth [21]. This theoretical understanding forms the basis for the empirical analysis conducted in this study, aiming to provide actionable insights for digital marketers.

Identification of Gaps in the Existing Literature

Despite the extensive research on engagement metrics and conversion, several gaps remain in the existing literature. First, while many studies have focused on specific industries or marketing channels, there is a lack of comprehensive analyses that compare the impact of engagement metrics across different contexts. This limitation makes it difficult to generalize findings and develop universally applicable strategies.

Additionally, most existing studies primarily utilize traditional statistical methods such as logistic regression and decision trees. While these methods provide valuable insights, they may not fully capture the complexity and interactions between multiple engagement metrics. Advanced machine learning techniques, such as random forests and gradient boosting machines, offer the potential to uncover more nuanced relationships and improve predictive accuracy, yet their application in this context remains underexplored.

Furthermore, there is limited research that integrates multiple engagement metrics to develop a holistic understanding of customer behavior. Many studies tend to analyze metrics in isolation, which can overlook the interplay between different forms of engagement. A more integrated approach that considers the combined effect of various metrics could provide deeper insights and more robust predictions.

This study aims to address these gaps by employing a comprehensive analytical approach that integrates logistic regression, decision trees, and random forests. By analyzing the impact of multiple engagement metrics across various digital marketing contexts, this research seeks to provide a more holistic understanding of the factors driving customer conversion. The findings will contribute to the existing body of knowledge and offer practical recommendations for optimizing digital marketing strategies.

Logistic Regression Formula

Logistic regression is a widely used statistical method for modeling binary outcomes, such as customer conversion (converted or not converted). Unlike linear regression, which predicts a continuous outcome, logistic regression is designed to predict the probability of a categorical dependent variable. The logistic regression model calculates the log-odds of the probability of an event occurring, which is then transformed into a probability value between 0 and 1 using the logistic function. The formula for logistic regression is:

$$P = \frac{1}{1+e^{-z}} z = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n \quad (1)$$

Decision Tree Algorithm

Decision trees are a type of supervised learning algorithm used for both classification and regression tasks. They work by recursively splitting the dataset into subsets based on the value of a chosen feature. Each node in the tree represents a feature, each branch represents a decision rule, and each leaf node represents an outcome (e.g., conversion or non-conversion). The goal of a decision tree is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.

The decision tree algorithm begins with the entire dataset as the root. The algorithm then selects the feature that best splits the data into distinct classes. This process, known as "splitting," is based on criteria such as Gini impurity or information gain. The selected feature creates branches, and the process is repeated for each branch, further splitting the data until the algorithm reaches a stopping criterion (e.g., maximum depth, minimum samples per leaf, or no further information gain). The resulting model is a tree where the path from the root to a leaf node represents a series of decisions leading to a classification or prediction.

Random Forest Algorithm

Random forests are an ensemble learning method that combines multiple decision trees to improve the accuracy and robustness of predictions. The fundamental idea is to create a "forest" of decision trees, each trained on a random subset of the data and features. By aggregating the predictions from multiple trees, the random forest algorithm reduces the variance and mitigates the risk of overfitting, which is a common issue with individual decision trees.

The random forest algorithm involves several key steps. First, it randomly selects a subset of the training data with replacement (bootstrapping) to create multiple subsets. For each subset, a decision tree is grown by randomly selecting a subset of features at each split. This randomness ensures that the trees are diverse and capture different patterns in the data. Once all trees are trained, the random forest makes predictions by averaging the predictions of individual trees (for regression) or using majority voting (for classification). Additionally, random forests provide an estimate of feature importance, indicating how much each feature contributes to the prediction accuracy. This is calculated by measuring the decrease in prediction accuracy when a feature is randomly permuted.

By leveraging logistic regression, decision trees, and random forests, this study aims to provide a comprehensive analysis of the impact of engagement metrics on customer conversion. Each method offers unique strengths: logistic regression for its interpretability, decision trees for their simplicity and visual representation, and random forests for their accuracy and robustness.

Methodology

The research method for this study consists of several steps to ensure a comprehensive and accurate analysis. The flowchart in figure 1 outlines the detailed steps of the research method.

Figure 1 Research Method Flowchart

Data Collection

The dataset utilized in this study is titled the "Predict Conversion in Digital Marketing Dataset." This dataset is specifically designed to offer a comprehensive view of customer interactions with various digital marketing campaigns. It includes detailed records of demographic information, marketing-related variables, engagement metrics, and historical purchasing behaviors of customers. The dataset serves as a robust foundation for analyzing how different engagement metrics influence the likelihood of customer conversion, which is the primary focus of this research.

The dataset consists of 8,000 entries, each representing a unique customer. There are 20 columns in total, capturing a broad spectrum of variables essential for a thorough analysis. The data is complete, with no missing values in any of the columns, ensuring the reliability of the analysis. This extensive dataset allows for a detailed exploration of the factors that drive conversion in digital marketing, making it particularly well-suited for predictive modeling using techniques such as logistic regression, decision trees, and random forests.

The features in the dataset can be broadly categorized into four groups: demographic information, marketing-specific variables, engagement metrics, and historical data, shown in Table 1.

Table 1. Features of Dataset

Category	Feature	Description	Data Type
Demographic Information	CustomerID	A unique identifier for each customer	int64
	Age	The age of the customer	int64
	Gender	The gender of the customer	object
	Income	The annual income of the customer in USD	int64
Marketing-Specific Variables	CampaignChannel	The channel through which the marketing campaign was delivered (e.g., Email, SocialMedia, SEO, PPC, Referral)	object
	CampaignType	The type of marketing campaign (e.g., Awareness, Consideration, Conversion, Retention)	object

Category	Feature	Description	Data Type
	AdSpend	The amount spent on the marketing campaign in USD	float64
	ClickThroughRate	The rate at which customers clicked on the marketing content	float64
	ConversionRate	The rate at which clicks converted to desired actions, such as purchases	float64
Engagement Metrics	WebsiteVisits	The number of times a customer visited the website	int64
	PagesPerVisit	The average number of pages visited per session	float64
	TimeOnSite	The average time spent on the website per visit, measured in minutes	float64
	SocialShares	The number of times the marketing content was shared on social media	int64
	EmailOpens	The number of times marketing emails were opened by the customer	int64
	EmailClicks	The number of times links within marketing emails were clicked by the customer	int64
Historical Data	PreviousPurchases	The number of previous purchases made by the customer	int64
	LoyaltyPoints	The number of loyalty points accumulated by the customer	int64

This table categorizes and describes each feature within the dataset, providing a clear and concise overview of the variables used in the study. The target variable in this dataset is Conversion (int64), a binary indicator of whether the customer converted (1) or did not convert (0). This variable is the focal point of the predictive modeling efforts in this study, as the primary goal is to identify which features most strongly influence conversion outcomes.

This well-rounded dataset provides a rich source of information for understanding the dynamics of customer behavior in response to digital marketing campaigns. The subsequent analysis leverages this data to uncover insights that can inform more effective marketing strategies, thereby enhancing conversion rates.

Data Preprocessing

Effective data preprocessing is a critical step in ensuring that the dataset is clean, consistent, and suitable for modeling. The data preprocessing phase for the "Predict Conversion in Digital Marketing Dataset" involved handling missing values and outliers, followed by the normalization and standardization of key engagement metrics to prepare the data for analysis.

The initial step in data preprocessing involved checking for missing values across all columns of the dataset. The results indicated that there were no missing values in any of the columns, confirming the dataset's completeness and integrity. This absence of missing data simplified the preprocessing phase, allowing the focus to shift toward the detection and treatment of outliers.

Outliers can significantly skew the results of statistical analyses and machine learning models, so identifying and addressing them is crucial. Two methods were employed to detect and manage outliers: the Z-score method and the Interquartile Range (IQR) method. The Z-score method identified outliers by flagging data points with a Z-score greater than 3 or less than -3, while the IQR method considered data points outside the range of Q1 - 1.5IQR to Q3 + 1.5IQR as outliers. The Z-score method detected no significant outliers, leaving the dataset intact with 8,000 entries. However, the IQR method identified and removed outliers, resulting in a reduced dataset with 7,012 entries.

With the dataset cleaned of outliers, the next step was to normalize and standardize the engagement metrics to ensure they were on comparable scales. This is particularly important for machine learning models that are sensitive to the magnitude of input features.

Normalization was applied to rescale the data to a range between 0 and 1. This technique is especially useful when the goal is to compare features that have different units or scales. For example, the WebsiteVisits feature was normalized to reflect values between 0 and 1, making it easier to compare with other metrics such as PagesPerVisit and TimeOnSite.

Standardization, on the other hand, was used to transform the data to have a mean of 0 and a standard deviation of 1. This method is beneficial when the features follow a normal distribution or when the objective is to apply algorithms that assume standardized data, such as logistic regression. After standardization, features like EmailClicks and SocialShares were centered around 0, with their variance standardized, ensuring that each feature contributed equally to the model.

Both normalization and standardization were applied to the engagement metrics, and the resulting transformed datasets were reviewed to ensure that the scaling was performed correctly. The first few rows of the normalized and standardized datasets revealed that the transformation processes were successful, with the metrics appropriately rescaled for further analysis.

This preprocessing stage set the foundation for accurate and reliable predictive modeling by ensuring that the dataset was free from biases introduced by missing values, outliers, or scale discrepancies. As a result, the models developed in subsequent steps are based on a well-prepared dataset, optimizing their performance and predictive power.

Exploratory Data Analysis (EDA)

EDA began with a comprehensive examination of the dataset's key variables using descriptive statistics. This analysis provided valuable insights into the distribution, central tendency, and variability of the dataset. The dataset contains 8,000 entries, with each entry representing a unique customer. Descriptive statistics reveal the mean, standard deviation, minimum, maximum, and quartile values for each variable, offering an overview of the dataset's characteristics.

For instance, the average age of customers is approximately 43.63 years, with a standard deviation of 14.90 years, indicating a moderately diverse age range among the customer base. Income varies significantly across the dataset, with an average annual income of $84,664 and a standard deviation of $37,580, suggesting considerable economic diversity among customers. The average amount spent on marketing campaigns (AdSpend) is approximately $5,001, with a maximum of nearly $10,000, reflecting a wide range of marketing investment levels. Engagement metrics such as WebsiteVisits, PagesPerVisit, and TimeOnSite show variability, with mean values of 24.75 visits, 5.55 pages per visit, and 7.73 minutes on site, respectively. These metrics highlight the differences in how customers interact with digital content.

The target variable, Conversion, indicates that 87.65% of customers converted, with a standard deviation of 0.33. This high conversion rate suggests that the dataset is somewhat imbalanced, with a majority of customers successfully completing the desired action. Understanding these statistics is crucial as they inform the subsequent modeling process by identifying potential biases or areas that may require further preprocessing.

To further explore the dataset, various visualizations were employed, including histograms, box plots, scatter plots, and a correlation matrix. These visual tools provided deeper insights into the distribution and relationships between the variables, particularly the engagement metrics and their connection to customer conversion.

Histograms for each engagement metric were generated to visualize their distribution across the customer base. For example, the histogram for WebsiteVisits in Figure 2 shows a skewed distribution, with most customers visiting the website fewer than 50 times, yet a small group of customers exhibiting much higher visit frequencies. Box plots were used alongside histograms to identify any potential outliers and to better understand the spread of data within each metric, shown in Figure 3. These visualizations revealed that while some metrics, like EmailOpens, follow a more uniform distribution, others, like SocialShares, have a wider range with significant variation.

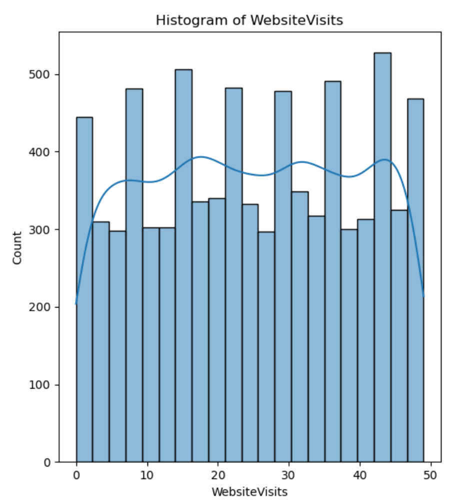

Figure 2 Histogram of Website Visits

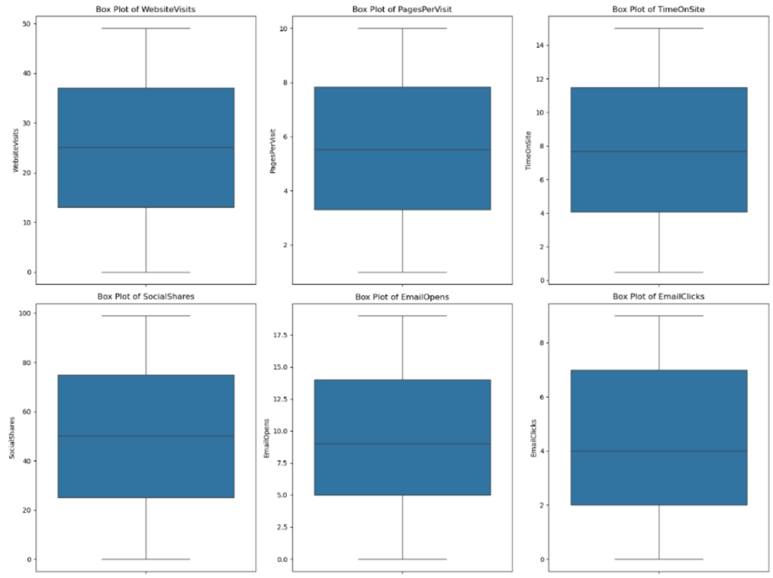

Figure 3 BoxPlot of Dataset Features

A correlation matrix was also computed and visualized to quantify the strength of relationships between the features. The correlation matrix in Figure 4 revealed that metrics such as PagesPerVisit and TimeOnSite have moderate positive correlations with each other, which is expected as customers who spend more time on the website are likely to visit more pages.

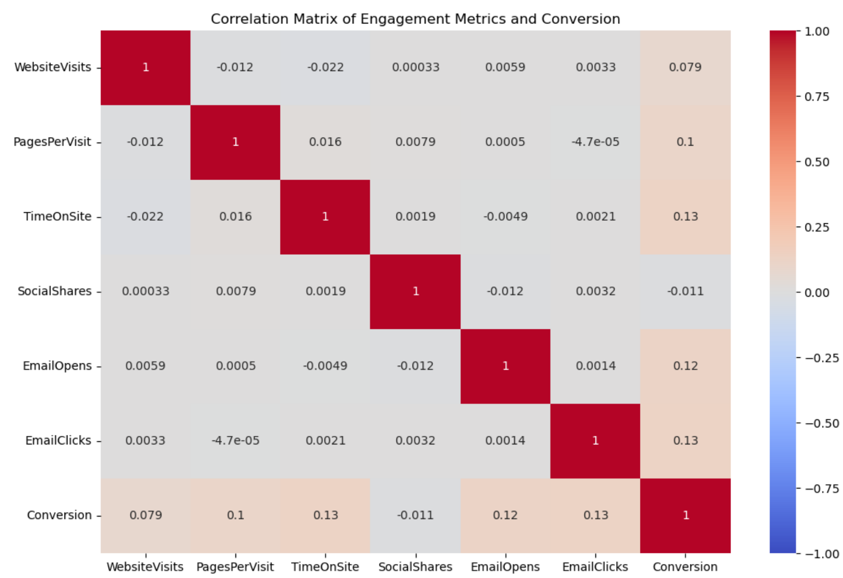

Figure 4 Correlation Matrix

The correlation matrix represents the relationships between various engagement metrics and the target variable, Conversion. Each cell in the matrix shows the correlation coefficient between two variables, with values ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). The color gradient in the matrix helps to quickly identify the strength and direction of these correlations.

Feature Selection

Feature selection is a crucial step in the predictive modeling process, as it involves identifying the most relevant variables that contribute significantly to predicting the target outcome—in this case, customer conversion. The goal is to refine the model by focusing on the most impactful features, thereby enhancing model performance and interpretability. This section outlines the methods used for feature selection, including correlation analysis and feature importance techniques such as Mutual Information and ANOVA F-tests.

Correlation analysis was the first step in assessing the relationships between the engagement metrics and the target variable, Conversion. The correlation matrix provided a visual and numerical summary of these relationships. Metrics such as TimeOnSite, EmailClicks, PagesPerVisit, and EmailOpens showed positive correlations with Conversion, with correlation coefficients ranging from 0.1 to 0.13. These metrics were identified as potentially important predictors, as they exhibited a moderate positive relationship with conversion rates. On the other hand, metrics like SocialShares showed near-zero correlation with Conversion, suggesting that they may not be significant predictors in the context of this dataset.

While correlation analysis provides insights into linear relationships, it does not capture more complex interactions or account for the potential impact of multiple variables acting together. Therefore, additional techniques were employed to further evaluate the importance of each feature. To complement the findings from the correlation analysis, more advanced feature importance techniques were applied, specifically Mutual Information and ANOVA F-tests.

Mutual Information is a non-linear method that measures the dependency between each feature and the target variable. It captures more complex relationships that may not be evident through correlation analysis alone. The results from the Mutual Information analysis indicated that TimeOnSite, EmailClicks, and EmailOpens had higher mutual information scores, suggesting that these features contain valuable information for predicting conversion.

ANOVA F-tests were conducted to statistically quantify the significance of each feature in explaining the variance in the conversion outcome. The F-test results reinforced the importance of several key metrics. TimeOnSite and EmailClicks both had high F-values of approximately 136, with extremely low p-values (close to zero), indicating their strong significance in predicting conversion. EmailOpens also had a high F-value of 126, confirming its relevance as a predictor. PagesPerVisit and WebsiteVisits, while still significant, had lower F-values compared to the email-related metrics and TimeOnSite, indicating a relatively weaker but still important contribution to the prediction model. SocialShares, with a low F-value of 1.048 and a p-value greater than 0.05, was deemed statistically insignificant in this context, aligning with the earlier correlation analysis.

These feature importance techniques collectively provided a robust basis for selecting the most relevant variables for the predictive models. The combination of linear correlation, non-linear dependency analysis through Mutual Information, and statistical significance testing via ANOVA allowed for a comprehensive evaluation of feature importance. As a result, the features TimeOnSite, EmailClicks, EmailOpens, PagesPerVisit, and WebsiteVisits were prioritized for inclusion in the subsequent modeling process, while SocialShares was excluded due to its lack of predictive power. This careful selection of features aims to enhance the efficiency and accuracy of the models developed in the next stages of the analysis.

Predictive Modeling

Model Formulation and Assumptions: Logistic regression is a statistical method used to model a binary outcome, such as whether a customer will convert (1) or not convert (0). The model estimates the probability of the target event occurring based on the input features. The logistic regression model assumes a linear relationship between the log-odds of the dependent variable and the independent variables. This assumption is crucial as it influences how the model fits the data and predicts outcomes.

In this study, logistic regression was applied to predict customer conversion using six key engagement metrics: WebsiteVisits, PagesPerVisit, TimeOnSite, SocialShares, EmailOpens, and EmailClicks. The model coefficients indicate the direction and magnitude of the relationship between each feature and the log-odds of conversion. For instance, positive coefficients suggest that an increase in the corresponding feature leads to higher odds of conversion, while negative coefficients imply the opposite.

The logistic regression model produced the following coefficients: WebsiteVisits (0.019), PagesPerVisit (0.126), TimeOnSite (0.102), SocialShares (-0.002), EmailOpens (0.081), and EmailClicks (0.157). These coefficients suggest that EmailClicks, PagesPerVisit, and TimeOnSite have the most substantial positive impact on conversion likelihood. Specifically, a one-unit increase in EmailClicks is associated with a 15.7% increase in the odds of conversion, making it the most influential feature. In contrast, SocialShares has a negligible and slightly negative effect, indicating it may not be a significant predictor of conversion in this context.

The model's performance was evaluated using metrics such as precision, recall, and F1-score. The classification report indicates an overall accuracy of 88%, with a particularly high recall (100%) for the conversion class (1). However, the model struggled with the non-conversion class (0), achieving only 3% recall, which reflects its limitation in distinguishing between converters and non-converters. This is further illustrated by the confusion matrix, where most of the actual non-conversions were misclassified as conversions.

A decision tree is a non-parametric model that predicts the target variable by learning simple decision rules inferred from the data features. The tree is constructed by recursively splitting the dataset into subsets based on the value of an input feature, which best separates the target classes at each node. The decision tree's structure makes it highly interpretable, as each path from the root to a leaf node represents a specific set of conditions leading to a prediction.

In this study, the decision tree model was built using the same six engagement metrics. The decision rules generated by the tree prioritize features like EmailClicks and PagesPerVisit at the top nodes, reflecting their importance in predicting conversion. The tree model's ability to capture non-linear relationships between features and the target variable provides a complementary approach to logistic regression.

The decision tree structure was visualized to illustrate the hierarchy of decision rules. The tree's top splits highlight the key role of EmailClicks in the conversion process, consistent with the logistic regression findings. However, the decision tree model exhibited lower overall accuracy (80%) compared to logistic regression. The classification report and confusion matrix revealed a relatively poor performance in predicting the non-conversion class, with precision and recall of 22% and 27%, respectively. The model also misclassified a significant number of non-converters as converters, indicating a tendency to overfit to the majority class (converters).

Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions. The key advantage of random forests lies in their ability to reduce overfitting, a common issue with individual decision trees, by averaging the predictions of multiple trees. Random forests also provide a measure of feature importance, which quantifies the contribution of each feature to the model's predictions.

In this analysis, the Random Forest model was trained using the six engagement metrics, and feature importance scores were calculated. The results indicated that PagesPerVisit (22.2%) and TimeOnSite (21.9%) were the most important features, followed by SocialShares and WebsiteVisits. Interestingly, despite EmailClicks having the highest coefficient in logistic regression, it was ranked lower in importance within the random forest model, highlighting the differences in how these models evaluate feature relevance.

The feature importance rankings were visualized to provide a clear understanding of the contribution of each metric. The bar plot showed that PagesPerVisit and TimeOnSite were the dominant features in predicting conversion, aligning with the findings from the logistic regression analysis. However, the random forest model produced slightly different rankings compared to the other models, which underscores the value of using multiple methods to gain a comprehensive understanding of feature importance.

The random forest model achieved an accuracy of 87%, similar to logistic regression but with better handling of the non-conversion class. The model's classification report revealed a precision of 40% and recall of 12% for the non-conversion class, indicating a slight improvement over the decision tree model. However, it still showed a strong bias towards predicting conversions, as evidenced by the confusion matrix.

Model Evaluation

To evaluate the performance of the predictive models, the dataset was first split into training and testing sets. This approach ensures that the models are evaluated on unseen data, providing a more realistic assessment of their predictive capabilities. Specifically, 80% of the dataset was allocated to the training set, which was used to train the models, while the remaining 20% was reserved as the testing set to evaluate the models' performance. This standard data splitting strategy helps to mitigate overfitting by ensuring that the models do not simply memorize the training data but instead generalize well to new, unseen examples.

To further enhance the reliability of the model evaluation, a 5-fold cross-validation process was employed. Cross-validation involves dividing the training data into five subsets, or "folds." The model is trained on four of these folds and validated on the remaining fold, a process that is repeated five times with a different fold used for validation each time. This technique ensures that every data point in the training set has the opportunity to be used for both training and validation, leading to a more robust estimate of the model's performance. The mean ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) score from these cross-validation runs was used as a key metric to assess each model's ability to distinguish between classes.

The performance of the models was evaluated using a range of metrics to provide a comprehensive assessment of their strengths and weaknesses. The primary metrics used were accuracy, precision, recall, F1-score, and ROC-AUC. Each metric offers different insights into the model's performance. Accuracy measures the proportion of correct predictions out of all predictions made, providing a general sense of the model's performance. Precision assesses the proportion of true positive predictions among all positive predictions, indicating how often the model correctly identifies conversions. Recall (sensitivity) measures the proportion of actual positive cases that were correctly identified, reflecting the model's ability to capture all potential conversions. F1-Score is the harmonic mean of precision and recall, offering a balanced metric that accounts for both false positives and false negatives. ROC-AUC provides a measure of the model's ability to distinguish between the two classes (conversion and non-conversion) across different threshold settings, with higher values indicating better performance.

The logistic regression model achieved an accuracy of 88.1% on the testing set, with a high recall of 99.8% for the conversion class. However, it exhibited a significant imbalance in performance, with a precision of only 67% for the non-conversion class, as reflected in the confusion matrix where most non-converters were misclassified. The mean ROC-AUC score across the cross-validation folds was 0.7072, indicating a moderate ability to discriminate between converters and non-converters.

The decision tree model, while interpretable, underperformed compared to logistic regression, with an accuracy of 79.6% and a mean ROC-AUC of 0.5751. The model's recall for the conversion class was 86.8%, but it struggled with precision for the non-conversion class, which was reflected in the low F1-score of 24% for that class. The confusion matrix highlighted the model's tendency to misclassify non-converters as converters, indicating potential overfitting to the majority class in the training data.

The random forest model provided a more balanced performance, achieving an accuracy of 87.1% and a mean ROC-AUC score of 0.6979. The model excelled in recall for the conversion class (97.5%) and maintained a reasonable precision (88.9%), resulting in a robust F1-score of 93.0%. However, like the other models, it still faced challenges in predicting the non-conversion class, with only 40% precision and 12% recall for non-converters. The random forest's ensemble approach helped reduce overfitting compared to the single decision tree, but the imbalance in class prediction remained a concern.

Results and Discussion

Model Performance

The performance of the three predictive models—Logistic Regression, Decision Trees, and Random Forests—was evaluated using key metrics: accuracy, precision, recall, F1-score, and ROC-AUC. These metrics provide a comprehensive overview of each model's strengths and limitations, allowing for a detailed comparison of their effectiveness in predicting customer conversion.

Comparison of Model Accuracy, Precision, Recall, F1-score, and ROC-AUC: The logistic regression model achieved the highest accuracy at 88.1%, with a particularly strong recall of 99.8% for the conversion class. However, its performance was less balanced, as indicated by its lower precision for the non-conversion class, which resulted in a weaker F1-score for this class. The model's ROC-AUC score of 0.7072 suggests that it has a moderate ability to distinguish between converters and non-converters, though it is biased towards predicting conversions.

The decision tree model, while more interpretable, underperformed relative to logistic regression, with an accuracy of 79.6% and a lower ROC-AUC of 0.5751. This model struggled particularly with the non-conversion class, as evidenced by its low precision and recall, resulting in a poor F1-score. The decision tree's simplicity makes it prone to overfitting, especially given the imbalanced nature of the dataset, where the majority class (conversions) dominated the predictions.

The random forest model offered a more balanced performance, achieving an accuracy of 87.1% and a ROC-AUC score of 0.6979. It managed to improve recall and precision for the conversion class compared to the decision tree, while still facing challenges in accurately predicting the non-conversion class. The ensemble nature of the random forest, which combines multiple decision trees, helped to mitigate some of the overfitting observed in the single decision tree model, but the imbalance in class prediction remained an issue. The table 2 summarizes the performance metrics for each model. Logistic regression demonstrated the highest accuracy and recall for the conversion class, but its low recall for the non-conversion class significantly impacted its overall performance.

Table 2. Model Performance Metrics

Metric	Logistic Regression	Decision Tree	Random Forest
Accuracy	88.1%	79.6%	87.1%
Precision (0)	67.0%	22.0%	40.0%
Precision (1)	88.2%	89.6%	88.9%
Recall (0)	3.0%	27.0%	12.0%
Recall (1)	99.8%	86.8%	97.5%
F1-Score (0)	6.0%	24.0%	18.0%
F1-Score (1)	93.6%	88.2%	93.0%

Metric	Logistic Regression	Decision Tree	Random Forest
ROC-AUC	0.7072	0.5751	0.6979

The decision tree, while interpretable, struggled with precision and recall, leading to the lowest ROC-AUC score. The random forest model provided a balanced approach, offering competitive accuracy and ROC-AUC scores while also improving on the decision tree's weaknesses.

Feature Importance

Understanding which features most significantly influence customer conversion is crucial for optimizing digital marketing strategies. The Random Forest model, with its ability to handle many features and measure their importance, provides valuable insights into which engagement metrics are the most predictive of conversion outcomes. By averaging the results across multiple decision trees, the Random Forest model ranks features based on their contribution to reducing impurity (Gini impurity or information gain) in the forest.

The feature importance analysis from the Random Forest model revealed that PagesPerVisit and TimeOnSite were the two most influential metrics, with importance scores of 22.2% and 21.9%, respectively. These metrics are closely related, as both capture aspects of customer engagement on the website. PagesPerVisit reflects how extensively a customer explores the site, while TimeOnSite indicates the duration of this engagement. The high importance of these features suggests that customers who spend more time on the website and view more pages are more likely to convert, highlighting the critical role of sustained and deep engagement in driving conversions.

Other features such as SocialShares and WebsiteVisits also contributed to the model’s predictions, though to a lesser extent, with importance scores of 16.2% and 15.9%, respectively. Interestingly, EmailOpens and EmailClicks, despite showing strong predictive power in the logistic regression model, were ranked lower in the Random Forest analysis, with importance scores of 13.1% and 10.6%. This discrepancy underscores the value of using multiple modeling approaches to capture different aspects of feature relevance.

The relatively high importance of SocialShares indicates that, while not the top predictor, social media engagement still plays a meaningful role in the conversion process. This suggests that customers who share content are somewhat more likely to convert, potentially due to the influence of social proof or deeper brand interaction.

To provide a clearer understanding of the relative importance of each feature, the feature importance rankings were visualized using a bar plot (see figure 5). This visualization highlights the dominance of PagesPerVisit and TimeOnSite in driving conversions, with both metrics clearly standing out from the rest. The bar plot also shows a gradual decline in importance from SocialShares and WebsiteVisits to EmailOpens and EmailClicks, reflecting the decreasing but still significant impact of these features.

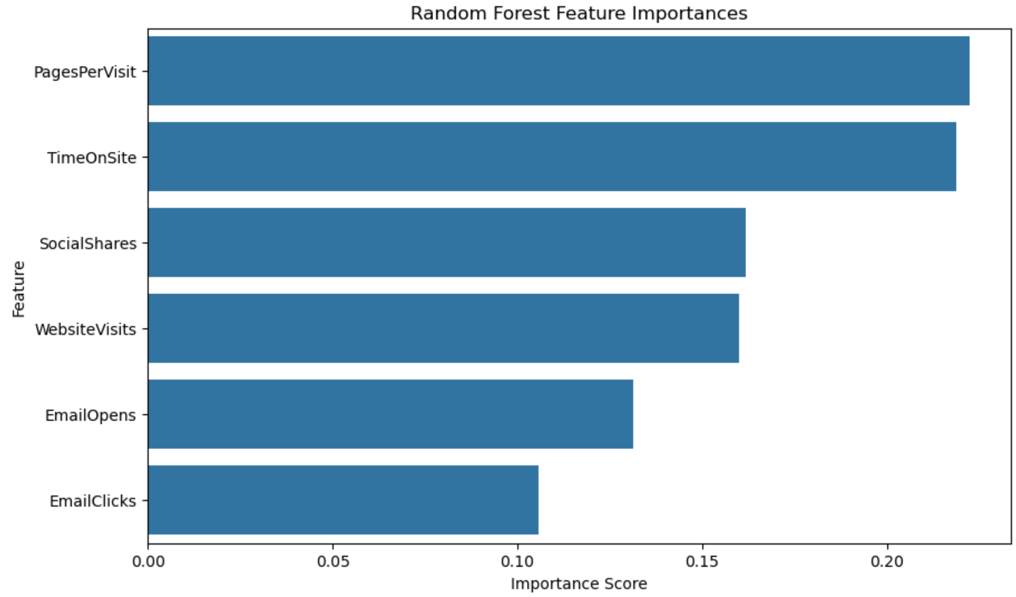

Figure 5 Feature Importance Score

The visualization of feature importance not only helps in interpreting the model's predictions but also informs marketing strategies. For example, the strong influence of PagesPerVisit and TimeOnSite suggests that optimizing website content to encourage deeper exploration and longer visits could be a key tactic for increasing conversion rates. Similarly, while EmailOpens and EmailClicks were less influential in the Random Forest model, their inclusion still supports the importance of targeted email campaigns as part of a broader engagement strategy.

This feature importance analysis reinforces the need for a multi-faceted approach in digital marketing, where both on-site engagement and off-site interactions like email and social media are strategically leveraged to drive conversions. The insights gained from the Random Forest model provide a robust foundation for optimizing these strategies, ensuring that resources are allocated to the most impactful areas of customer engagement.

Interpretation of Results and Their Implications

The results of this study provide significant insights into the factors that influence customer conversion in digital marketing contexts. The logistic regression, decision tree, and random forest models all highlighted the importance of customer engagement metrics, though with varying degrees of emphasis on different features. PagesPerVisit and TimeOnSite consistently emerged as critical predictors, suggesting that the depth and duration of customer interaction with a website are crucial determinants of conversion likelihood. These findings imply that digital marketing efforts should focus on enhancing user experience to increase these metrics, thereby improving conversion rates.

Moreover, the strong performance of EmailClicks in the logistic regression model indicates that direct engagement with email content remains a powerful tool for conversion. This suggests that personalized and targeted email campaigns could play a key role in driving customer actions. The varying importance of features across models underscores the need for a multi-faceted approach in digital marketing strategies, where different aspects of customer behavior are leveraged to maximize conversion outcomes.

Practical Insights for Digital Marketers

The practical implications of these findings for digital marketers are substantial. First, the emphasis on PagesPerVisit and TimeOnSite suggests that content strategies should prioritize quality and relevance, encouraging users to spend more time on the site and explore more pages. This could involve optimizing website design for better navigation, creating engaging content that resonates with users, and implementing recommendations that guide users through a more extensive exploration of the site.

Second, the importance of EmailClicks points to the effectiveness of email marketing campaigns. Marketers should focus on crafting compelling email content that drives clicks, such as personalized offers, clear calls to action, and engaging subject lines. Integrating email campaigns with broader digital marketing efforts, such as retargeting ads and social media promotions, could further enhance their impact on conversion.

Finally, the varying influence of SocialShares suggests that while social media engagement is beneficial, it may not be as critical as direct on-site interactions or email engagement. However, it still plays a role in overall brand visibility and customer engagement, indicating that social media should be part of a comprehensive marketing strategy.

Comparison with Previous Research Findings

The findings of this study align with previous research that highlights the importance of customer engagement in predicting conversion. Studies have consistently shown that metrics related to website interaction, such as time spent on site and the number of pages visited, are strong predictors of customer behavior (e.g., Huang & Benyoucef, 2017). The high importance of EmailClicks also resonates with existing literature, which underscores the effectiveness of email marketing in driving conversions (Jenkins & Molesworth, 2019).

However, this study diverges from some earlier research in its finding of relatively low importance for SocialShares. While previous studies have suggested a stronger link between social media activity and conversion (Goh, Heng, & Lin, 2013), the current analysis suggests that social media may play a more supportive role rather than being a primary driver of conversion. This discrepancy could be due to differences in dataset characteristics, the nature of the campaigns analyzed, or evolving trends in digital consumer behavior.

Limitations of the Study and Suggestions for Future Research

Despite its contributions, this study has several limitations that should be addressed in future research. First, the dataset used is specific to a particular set of digital marketing campaigns, which may limit the generalizability of the findings. Future studies could explore a more diverse range of campaigns across different industries to validate the results and broaden their applicability.

Second, while the models used in this study provided valuable insights, they may not capture all the complexities of customer behavior. Advanced techniques such as deep learning or ensemble methods combining multiple algorithms could be explored to improve predictive accuracy and uncover more nuanced patterns.

Third, the study focused on a limited set of engagement metrics. Future research could incorporate additional variables, such as customer demographics, psychographics, or external factors like economic conditions, to provide a more comprehensive understanding of the factors influencing conversion.

Finally, the study's cross-sectional design limits its ability to capture the dynamic nature of customer behavior over time. Longitudinal studies tracking customer engagement and conversion across multiple touchpoints and time periods could offer deeper insights into the evolution of customer journeys and the long-term impact of digital marketing efforts.

Conclusion

This study set out to explore the key factors driving customer conversion in digital marketing by analyzing engagement metrics using three predictive modeling techniques: Logistic Regression, Decision Trees, and Random Forests. The analysis revealed that PagesPerVisit, TimeOnSite, and EmailClicks emerged as the strongest predictors of conversion across all models. These metrics, which capture the depth of customer interaction with digital content and their engagement with email campaigns, consistently demonstrated their importance in influencing conversion outcomes. The Random Forest model, in particular, highlighted PagesPerVisit and TimeOnSite as the most significant features, emphasizing the role of sustained and meaningful engagement in driving conversions.

In contrast, metrics like SocialShares, while still relevant, were found to have a lesser impact on conversion compared to direct on-site engagement and email interactions. These findings suggest that while social media activities contribute to overall marketing effectiveness, they may not be the primary drivers of conversion within the contexts examined in this study.

The practical implications of these findings for digital marketers are substantial. The identification of PagesPerVisit, TimeOnSite, and EmailClicks as key predictors underscores the importance of creating a website experience that encourages extensive exploration and engagement. Marketers should focus on optimizing website content and structure to retain users longer and encourage them to visit more pages, thereby increasing the likelihood of conversion. Additionally, the critical role of email engagement in conversion suggests that targeted, personalized email campaigns should be a central component of digital marketing strategies. These campaigns should be designed to prompt immediate action, such as clicking on links or exploring additional content, to maximize their effectiveness.

Moreover, the relative importance of these metrics over social media engagement indicates that while social media is an important tool for brand awareness and engagement, direct interactions with digital content on owned platforms (like websites and emails) may have a more immediate impact on conversion rates. This insight can guide marketers in prioritizing their efforts and resources, ensuring that the most impactful strategies are employed to drive customer conversion.

This study contributes to the existing body of knowledge in digital marketing analytics by providing empirical evidence on the relative importance of different engagement metrics in predicting customer conversion. While previous research has explored various aspects of digital marketing effectiveness, this study uniquely combines three powerful predictive models to offer a comprehensive analysis of how specific metrics influence conversion outcomes. The findings add depth to our understanding of digital customer behavior, particularly in how different forms of engagement—whether through website interaction or email engagement—affect conversion likelihood.

Furthermore, the study's use of multiple modeling techniques allows for a nuanced comparison of model performance and feature importance, highlighting the value of using diverse analytical approaches to uncover different facets of customer behavior. This contributes to the methodological discourse in data science, demonstrating the benefits of combining traditional statistical methods with more advanced machine learning techniques.

While this study provides valuable insights, it also opens up several avenues for future research. One limitation of the study is its focus on a specific set of digital marketing campaigns, which may limit the generalizability of the findings. Future research could expand on this work by analyzing a broader range of campaigns across different industries and regions to validate and extend the findings.

Additionally, this study primarily examined engagement metrics in isolation. Future studies could explore the interactions between these metrics and other variables, such as customer demographics, purchase history, or external factors like economic conditions, to build a more comprehensive model of customer conversion. Longitudinal studies that track customer behavior over time could also provide deeper insights into the dynamics of conversion, allowing researchers to understand how engagement evolves throughout the customer journey and its long-term impact on conversion rates.

Finally, the exploration of more sophisticated modeling techniques, such as deep learning or hybrid models that combine multiple algorithms, could further enhance predictive accuracy and uncover more complex patterns in customer behavior. These approaches could help in refining digital marketing strategies to better target potential customers and increase conversion rates in an increasingly competitive digital landscape.

Declarations

Author Contributions

Conceptualization: A.B.P., B.M.A., and A.A.; Methodology: B.M.A.; Software: A.B.P.; Validation: A.B.P., B.M.A., and A.A.; Formal Analysis: A.B.P., B.M.A., and A.A.; Investigation: A.B.P.; Resources: B.M.A.; Data Curation: B.M.A.; Writing—Original Draft Preparation: A.B.P., B.M.A., and A.A.; Writing—Review and Editing: B.M.A., A.B.P., and A.A.; Visualization: A.B.P. 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] R. Adepu and R. S. Rao, "Customer Engagement: The Next Frontierfor the Marketers," Mark. Brand. Res., 2020, doi: 10.33844/mbr.2020.60321.

[2] A. T. Nurlibna, "Analysis of the Influence of Advertising Creativity, Product Campaigns, and Brand Ambassador Credibility on Customer Loyalty for the Wearing Klamby Brand," J. Econ. Bus. Ubs, vol. 13, no. 2, pp. 438–450, 2024, doi: 10.52644/joeb.v13i2.1533.

[3] D. J. Petzer and E. v. Tonder, "Loyalty Intentions and Selected Relationship Quality Constructs," Int. J. Qual. Reliab. Manag., vol. 36, no. 4, pp. 601–619, 2019, doi: 10.1108/ijqrm-06-2018-0146.

[4] N. Barhemmati and A. Asghari, "Effects of Social Network Marketing (SNM) on Consumer Purchase Behavior Through Customer Engagement," J. Adv. Manag. Sci., pp. 307–311, 2015, doi: 10.12720/joams.3.4.307-311.

[5] N. A. Shofiya and I. Fachira, "Effects of Social Media Marketing Towards Probiotic Chicken Customers' Purchase Intention: Customer Engagement as a Mediator," Malays. J. Soc. Sci. Humanit. Mjssh, vol. 6, no. 8, pp. 518–531, 2021, doi: 10.47405/mjssh.v6i8.943.

[6] S. Logalakshmi, "CARVING a BRIGHTER PATH WITH SYNERGY OF DIGITAL MARKETING &Amp; AI," Int. J. Trendy Res. Eng. Technol., vol. 07, no. 05, pp. 18–24, 2023, doi: 10.54473/ijtret.2023.7505.

[7] F. L. Y. Aityassinec, M. M. Al-Ajlouni, and A. Mohammad, "The Effect of Digital Marketing Strategy on Customer and Organizational Outcomes," Mark. Manag. Innov., vol. 13, no. 4, pp. 45–54, 2022, doi: 10.21272/mmi.2022.4-05.

[8] M. Chinakidzwa and M. A. Phiri, "Impact of Digital Marketing Capabilities on Market Performance of Small to Medium Enterprise Agro-Processors in Harare, Zimbabwe," Verslas Teor. Ir Prakt., vol. 21, no. 2, pp. 746–757, 2020, doi: 10.3846/btp.2020.12149.

[9] O. D. A. Andrade, "The Power of Growth Hacking Strategies and the Exponential Growth of UBER," Int. J. Res. Mark. Manag. Sales, vol. 2, no. 1, pp. 16–24, 2020, doi: 10.33545/26633329.2020.v2.i1a.40.

[10] E. Bruce, S. Zhao, Y. Du, M. Yaqi, J. Amoah, and S. B. Egala, "The Effect of Digital Marketing Adoption on SMEs Sustainable Growth: Empirical Evidence From Ghana," Sustainability, vol. 15, no. 6, p. 4760, 2023, doi: 10.3390/su15064760.

[11] J. G. W. GUERRERO-VILLEGAS, "Digital Marketing Strategies and Acquisition of New Customers for Smes in Santa Elena Province," Russ. Law J., vol. 11, no. 9s, 2023, doi: 10.52783/rlj.v11i9s.1800.

[12] A. Risdwiyanto, "Sustainable Digital Marketing Strategy for Long-Term Growth of MSMEs," J. Contemp. Adm. Manag. Adman, vol. 1, no. 3, pp. 180–186, 2023, doi: 10.61100/adman.v1i3.70.

[13] Q. N. Wijayani, "The Potential for Tourism Development Based on Natural Wealth With a Digital Marketing Approach," pp. 362–372, 2023, doi: 10.2991/978-2-38476-118-0_41.
[14] R. A. Adeleye, "Digital Marketing Analytics: A Review of Strategies in the Age of Big Data and AI," World J. Adv. Res. Rev., vol. 21, no. 2, pp. 073–084, 2024, doi: 10.30574/wjarr.2024.21.2.0395.
[15] A. S. Krishen, Y. K. Dwivedi, N. Bindu, and K. S. Kumar, "A Broad Overview of Interactive Digital Marketing: A Bibliometric Network Analysis," J. Bus. Res., vol. 131, pp. 183–195, 2021, doi: 10.1016/j.jbusres.2021.03.061.
[16] L. Kapustina, O. Gaiterova, N. Izakova, and M. Lazukov, "Digital Marketing Communications: Selection Criteria," Kne Soc. Sci., 2021, doi: 10.18502/kss.v5i2.8351.
[17] B. Noveriyanto and S. E. Adawiyah, "Digital Integrated Marketing Communications (Dimc) Activities of Digital Products Financial Technology (Fintech) 'Alami,'" Profetik J. Komun., vol. 14, no. 1, p. 60, 2021, doi: 10.14421/pjk.v14i1.2017.
[18] H. Wijaya, "Impact of Digital Marketing and Intellectual Capital on Business Performance (Case Study of SMEs in Depok City, West Java)," Best J. Adm. Manag., vol. 2, no. 4, pp. 183–189, 2024, doi: 10.56403/bejam.v2i4.186.
[19] R. G. Duffett, "Customer Engagement and Intention to Purchase Attitudes of Generation Z Consumers Toward Emojis in Digital Marketing Communications," Young Consum. Insight Ideas Responsible Mark., vol. 25, no. 5, pp. 607–624, 2024, doi: 10.1108/yc-08-2023-1817.
[20] Madhukumar. B. Swapna Datta Khan Maria Antony Raj M., "Significant Role of Digital Marketing Strategies in Driving Business Growth, Success and Customer Experience," Jier, vol. 4, no. 2, 2024, doi: 10.52783/jier.v4i2.837.
[21] A. Saputra, "The Importance of Digital Marketing Integration in Strategic Management Planning," Action Res. Lit., vol. 8, no. 5, 2024, doi: 10.46799/arl.v8i5.340.